

# Diagnostic decision support for general practitioners

Citation for published version (APA):

Dupuits, F. M. H. M. (1997). *Diagnostic decision support for general practitioners*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.19971120fd>

## Document status and date:

Published: 01/01/1997

## DOI:

[10.26481/dis.19971120fd](https://doi.org/10.26481/dis.19971120fd)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# **Diagnostic Decision Support for General Practitioners**

ISBN 90-5681-021-9

© F.M.H.M. Dupuits, Maastricht, 1997

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system without written permission from the author.

Cover design: François M.H.M. Dupuits

Printed by: **Unigraphic**, Maastricht University

# **Diagnostic Decision Support for General Practitioners**

## **PROEFSCHRIFT**

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,  
op gezag van de Rector Magnificus, Prof.mr. M.J. Cohen,  
volgens het besluit van het College van Decanen,  
in het openbaar te verdedigen  
op donderdag 20 november 1997 om 12.00 uur

door

**François Marie Hubert Marcel Dupuits**

geboren op 13 december 1959 te Meerssen



**Promotores:**

Prof.dr.ir. A. Hasman  
Prof.dr. P. Pop

**Beoordelingscommissie:**

Prof.dr. C. Spreeuwenberg (voorzitter)  
Prof.dr. J.A. Knottnerus  
Dr.ir. P.P.C.A. Menheere  
Prof.dr. P.F. de Vries Robbé (Katholieke Universiteit Nijmegen)  
Dr. B.H.R. Wolffenbuttel

The research presented in this thesis was performed at the Department of Medical Informatics within an organizational affiliation with the Research Institute for Extramural and Transmural Health Care (Institute ExTra), a participant in The Netherlands School of Primary Care Research (Graduate School CaRe), which was acknowledged by the Royal Dutch Academy of Arts and Sciences (Dutch acronym: KNAW) in 1995.

The publication of this thesis was supported by grants from: Cap Gemini Nederland B.V. in Utrecht, Euroned Systems in Sittard, GlaxoWellcome in Zeist, Nourypharma B.V. in Oss, PharmaPartners B.V. in Oosterhout, and SMS Cendata in Nieuwegein. This support is gratefully acknowledged.

Thanks are expressed to all those involved in the research presented in this thesis.

Aan mijn ouders en  
mijn zus Monique

# Table of Contents

Chapter 1	Introduction	1
Chapter 2	Literature review	9
	General Practitioner Information Systems	10
	Decision Support for General Practitioners	14
	Medical Knowledge of Thyroid Disorders	20
Chapter 3	Computer-based assistance in family medicine	37
Chapter 4	Patients, methods, and techniques in the patient-oriented study of thyroid disorder prediction instruments	55
Chapter 5	Evaluation of a decision support system in a medical environment	67
Chapter 6	User satisfaction of general practitioners with HIOS+, a medical decision support system	83
Chapter 7	Evaluation of GPs' medical knowledge by a decision support system	93
Chapter 8	A rational request behaviour - The development of prediction instruments regarding thyroid function tests in primary care	103
	Prediction instruments for hyperthyroidism	104
	Prediction instruments for hypothyroidism	116
Chapter 9	Hyperthyroidism and hypothyroidism in outpatient and inpatient clinics	125
Chapter 10	Conclusions and general discussion	135
Chapter 11	Summary	147
Chapter 12	Samenvatting	155
Appendices to the Thesis		
List of Publications		
Curriculum Vitae		
List of Abbreviations and Acronyms		

## Chapter 1

### Introduction

## **Background**

The seventies and early eighties were characterized by a huge development in Information Technology (IT) which influenced all application areas (disciplines) in society. By the mideighties, computers had also found their way to the field of primary care in The Netherlands. Until then, Dutch General Practitioners (GPs) had made use of paper-based medical records and unstructured data. IT made it possible to structure medical data in computer-based medical records.

When the first personal computers became available, it became apparent that the fast development of IT caused problems: GPs had difficulty in getting a grip on the vast supply of and in learning to employ the new technologies. To solve these problems, requirements and wishes of GPs concerning the use of information systems needed to be investigated, and general system design guidelines should be specified by an independent national body. As a result, the Dutch College of General Practitioners stipulated guidelines for the design of medical information systems in a 1985 reference guide<sup>1</sup>. In this reference guide, a GP information system was defined in terms of a number of modules in which each module performs specific functions. The guide focused especially on the module for patient registration, but discussed medical modules for registering patient encounters, problem lists, and medication, and modules for research purposes, diagnostic archives, and communication only briefly.

At the same time, also the Department of Medical Informatics (MI) of the Maastricht University (formerly called the University of Limburg) initiated research into requirements and wishes of GPs with respect to information systems to be used in daily practice and for research purposes. In order to actually offer GPs various design alternatives of information systems, to familiarize GPs with applications, and to assess these applications in daily practice, the GP Information and Research System HIOS<sup>2</sup> was developed. The aim was to determine system functions needed by GPs in daily practice and research, and to develop a system architecture in which these functions were integrated in a userfriendly way. HIOS was given a modular structure, consisting of six modules of which the one for patient registration was based on the 1985 reference guide. HIOS eventually became a system with which GPs can not only register patient data and medical data but also can analyse data for research purposes. It was one of the first systems in The Netherlands to use classification codes and a problem-oriented way of registering data in order to enable computerized data analysis.

As the use of IT in primary care progressed, GPs gained access to a growing bulk of medical and other data. They increasingly experienced difficulty in mastering and

interpreting the available data. After the study of design requirements for GP information and research systems was ended in 1989, research was concentrated on IT applications which can support GPs in data interpretation and provide them with an easy access to knowledge bases. One of these applications is computer-based decision support. The initial aim of the research was to theoretically investigate the use of decision support in daily primary care. In 1989, a one-month survey<sup>3</sup> was held among 24 university-affiliated GPs to assess the attitude of GPs towards the possibilities of computer-based decision support. The following questions were investigated:

- *Do GPs have a need for decision support in daily practice?*
- *Which types of decision support are needed or preferred?*
- *In which medical domains is decision support thought to be especially useful?*

By means of a questionnaire, the GPs were asked about medical situations in which they would have liked to be supported by a Decision Support System (DSS). They only were allowed to describe situations which actually occurred in their general practices or health centres, and not theoretical ones. Prior to the survey, the GPs were thoroughly informed about the possibilities of active decision support in which a GP is advised (in the form of either an advice or a warning) by a DSS regarding actions to be taken, and about the possibilities of passive decision support in which a GP is presented textbook information. Reasons for distinguishing these two types of decision support were that they:

- require different technical realizations,
- differently influence the decision-making process of GPs,
- have a different impact on the use of a DSS by GPs, and
- indicate different user needs.

For determining and ranking medical domains, the GPs were asked to specify domains and their priority values. It was concluded from the survey that GPs indeed had a need for decision support by means of a DSS. Active as well as passive support were required in many instances. As for the medical domains, the GPs stated that they were most in need of active support concerning the diagnosis of thyroid disorders, the diagnosis and the treatment of hypertension, and causes of dizziness; and of passive support regarding normal ranges of laboratory test outcomes, characteristics of medication, and medical knowledge concerning the diagnosis of diseases on the basis of criteria.

### **IT in primary care decision support**

On the basis of the results of the survey, it was decided to study the feasibility of

computer-based decision support in primary care, i.e. to investigate whether and in which ways IT could be applied to support the decision process of GPs. In order to actually offer GPs various decision support alternatives, to familiarize GPs with applications, and to assess these applications in daily practice, a DSS called HIOS+<sup>4</sup> was developed by extending HIOS with facilities for passive as well as active decision support.

The IT research involved the development, the implementation, and the evaluation of DSSs for GPs as well as an investigation of the impact of decision support on primary care. The principal research questions were:

- *What is the state of the art of GP information systems and DSSs?*
- *Can the quality of registered diagnoses and the completeness of medical records be enhanced by DSSs?*
- *With which type of decision support - active or passive - provided by DSSs are GPs most satisfied?*

An additional principal research question dealt with decision support IT applied to one medical domain, namely thyroid disorders. The investigated question was:

- *What is the GPs' level of knowledge regarding the medical domain of thyroid disorders when compared to the knowledge stored in the models of a DSS?*

The research into this question constituted a natural transition to the next area of research.

### **Thyroid disorder decision making in primary care**

On the basis of the survey, thyroid disease was chosen as the medical domain for decision support. The thyroid gland<sup>5</sup> is situated in the front of the neck. The thyroid function is controlled by the Thyroid-Stimulating Hormone (TSH). TSH secretion is regulated through feedback mechanisms. The principal hormones secreted by the thyroid gland are Thyroxine (T<sub>4</sub>) and Triiodothyronine (T<sub>3</sub>). Disruptions of the secretions result in the signs, symptoms, and complications which are attended with the two main thyroid disease categories of hypothyroidism and hyperthyroidism. A reduced secretion of thyroid hormones is called hypothyroidism. It is indicated by coarse and sparse hair, a dry skin, poor tolerance of cold, a husky and slow voice, a poor memory, mental and physical slowing, and sometimes severe mental symptoms. An increased secretion of thyroid hormones is called hyperthyroidism. It leads to nervousness, weight loss, hyperphagia, heat intolerance, an increased pulse pressure, a fine tremor of the outstretched fingers, a warm and soft skin, tachycardia, and excess heat production resulting in sweating. It is often caused by MultiNodular Struma (MNS) or Graves' disease. These disorders are

relatively frequent (35% and 50% respectively). Whereas MNS often occurs at a later age, Graves' disease is frequent among the under forties. The clinical picture of MNS is less clear than that of Graves' disease, which makes MNS more difficult to diagnose. Hyperthyroidism can be treated by surgical intervention or with thyrostatic (inhibiting) drugs or radioactive iodine ( $^{131}\text{I}$ ) uptakes, and hypothyroidism with thyromimetic (supplementing) drugs.

In the survey, the GPs had indicated that they had difficulty in effectively detecting patients with hyperthyroidism or hypothyroidism. They had stressed the complexity of the clinical pictures of thyroid disorders, which range from various signs/symptoms in young to monosymptomatic disease patterns<sup>6</sup> in elderly people. Research, therefore, was focused on investigating and structuring the medical knowledge concerning thyroid disorders for decision support purposes.

This research coincided with the research performed at the Diagnostic Coordinating Centre (DCC) of the University Hospital Maastricht into the adequacy of diagnostic test requests of GPs. The DCC gives feedback to GPs about their diagnostic test request behaviour, i.e. their requests for diagnostic tests in the fields of e.g. haematology, biochemistry, electrocardiography, histology, and diagnostic imaging (radiology, endoscopy, etc.). The DCC aims at stimulating the rational diagnostic behaviour of GPs on the basis of work agreements created and/or agreed upon by GPs, standards formulated by the Dutch College of General Practitioners, and principles of medical decision making. These principles involve an estimation of the pretest probability of disease (diagnostic/therapeutic thresholds); a selection of diagnostic tests on the basis of aspects like discriminative values (e.g. sensitivities, specificities, likelihood ratios), complications (e.g. invasive tests), and costs; and consequences for disease management. The feedback to GPs about their diagnostic behaviour is given twice a year in structured reports and is based on clinical patient data. These reports among other things contain notes on unjustly requested tests. The notes refer to work agreements, standards, and decision-making principles. The feedback has positive results in that it leads to an improvement in the quality of test requests and a reduction in the number of unnecessary requests<sup>7,8</sup>. The same effects of feedback also were witnessed (on a temporary basis) in other studies in which paper-based or computer-based feedback was given in hospital care<sup>9-11</sup>. One of the diagnostic areas in which a more rational behaviour of GPs is stimulated is the diagnosis of thyroid disorders, i.e. requests of TSH tests (and, if necessary, Free  $\text{T}_4$  ( $\text{FT}_4$ ) tests) for patients with an increased pretest probability of hypothyroidism or hyperthyroidism. The DCC experienced that GPs request relatively



many TSH tests of which the majority (90% of the cases) has normal outcomes. GPs request a TSH test for check-up, screening, or exclusion reasons rather than to confirm a diagnosis. The tests, in other words, are requested for patients with a low pretest probability on the basis of signs/symptoms. In fact, most of the patients have only a few signs/symptoms or a combination of a few rather atypical signs/symptoms. In order to be able to give better advice to GPs in this matter, it is important to know the predictive values of combinations of signs/symptoms with regard to diagnosing hypothyroidism or hyperthyroidism. Only if a pretest probability exceeds a diagnostic threshold, a TSH test should be requested. This course of action should stimulate a rational diagnostic behaviour of GPs and result in a reduction in costs by the prevention of unnecessary TSH tests. These tests are relatively expensive in comparison with other tests at the disposal of GPs.

The DCC's thyroid disorder research was linked up with MI's research. The joined research aimed at providing GPs with (computer-based) prediction instruments with which they can determine the adequacy of TSH test requests when they intend to request such a test but before actually making a request. The hypothesis was that by having GPs use pretest probabilities of thyroid disorders (hyperthyroidism and hypothyroidism), which are estimated on the basis of signs/symptoms, in combination with diagnostic/therapeutic thresholds, a more rational diagnostic behaviour of GPs regarding hyperthyroidism and hypothyroidism could be obtained. The principal research question investigated in a prospective patient-oriented study was:

- *Can prediction instruments based on signs and symptoms be developed which indicate whether a TSH test is adequate?*

In daily practice, a GP should use a prediction instrument during a patient encounter when he/she considers requesting a TSH test for a patient and before actually ordering the test. By using one or more instruments, a GP should be able to confirm or exclude a suspicion of hyperthyroidism and/or hypothyroidism and, thus, to ascertain TSH test adequacy.

Since the number of patients found in the prospective study was lower than expected, additionally the following research question was investigated in a retrospective patient-oriented study:

- *How many patients with thyroid disease for whom GPs do not request a TSH test are detected in hospital; where are they detected; and why are they not detected in primary care?*

## The structure of the thesis

The structure of the thesis expresses the fact that the research covers two main areas, i.e. IT in primary care decision support and thyroid disorder decision making in primary care. In the general chapters 1, 2 and 10, both subjects are described. Chapters 3 and 5 through 7 deal with experiments of IT in medical decision support. Chapters 4, 8, and 9 cover the patient-oriented studies of thyroid disorder decision making.

**Chapter 2** contains the results of a literature review of both research areas and is divided into three parts regarding General Practitioner Information Systems, Decision Support for General Practitioners, and Medical Knowledge of Thyroid Disorders. **Chapter 3** deals with the software environment and the functionality of HIOS and several DSSs within HIOS+. In **chapter 4**, patients, methods, and techniques are described which were used to develop prediction instruments for thyroid disorders. In this chapter, also the quality of data acquired through questionnaires and the questionnaires themselves are dealt with. The impact of a goal-oriented reasoning system for GPs on the quality of recorded diagnoses and the completeness of medical records is covered in **chapter 5**. **Chapter 6** deals with user satisfaction aspects of this DSS. An educative DSS for the evaluation of the level of GPs' knowledge regarding thyroid disorders and its user evaluation are discussed in **chapter 7**. In **chapter 8**, results of the prospective patient-oriented study are described, including the prediction instruments for hyperthyroidism and hypothyroidism and their pretest probabilities, sensitivities, specificities, predictive values, and likelihood ratios. **Chapter 9** deals with the retrospective patient-oriented study of the detection of clinical hyperthyroid and clinical hypothyroid patients in hospital. **Chapter 10**, finally, contains conclusions and a general discussion of the research.

## References

1. The Adam Project Team. A general practitioners reference model - Global architecture definition. Brussels: European Community, 1990; AIM project 1023 (Deliverable 1.2).
2. Dupuits FMHM, Hasman A, Heumakers TAJ, Schoonbrood GGM. A research tool for general practitioners. *Int J Biomed Comp* 1987;21(3-4):275-286.
3. Dupuits FMHM, Hasman A. HIOS en besluitvormingsondersteuning. Readings in Medical Informatics - Proceedings van het Medisch Informatica Congres (MIC) 1989. Antwerpen: MIM/VMBI, 1989:257-269.
4. Dupuits FMHM, Hasman A. Designing HIOS+. In: Timmers T, Blum BI, eds. Proceedings of the IMIA Working Conference on Software Engineering in Medical Informatics (SEMI) 1992. Amsterdam: North-Holland Publishing Company, 1992:181-193.
5. Ganong WF. Review of medical physiology. 13th ed. Norwalk: Appleton & Lange, 1987:262-275.
6. Levy EG. Thyroid disease in the elderly. *Med Clin North Am*. 1991;75(1 - Special on Thyroid Diseases):151-167.

7. Winkens RAG, Pop P, Bugter-Maessen AMA, et al. Effects of routine individual feedback over nine years on general practitioners' requests for tests. *Br Med J* 1996;312:490.
8. Winkens RAG, Pop P, Bugter-Maessen AMA, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *The Lancet* 1995;345:498-502.
9. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for the outpatient diagnostic tests. *N Engl J Med* 1990;322:1499-1504.
10. Dowling PT, Alfonsi G, Brown MI, Culpepper L. An education program to reduce unnecessary laboratory tests by residents. *Acad Med* 1989;64:410-412.
11. McDonald CJ, Hui SL, Smith DM, et al. Reminders to physicians from an introspective computer medical record - A two-year randomized trial. *Ann Intern Med* 1984;100:130-138.

## **Chapter 2**

### **Literature review**

In this chapter, the evolution and the state of the art of General Practitioner (GP) Information Systems and Decision Support Systems (DSSs) as well as the use of medical knowledge of thyroid disorders by GPs are described.

### **General Practitioner Information Systems**

#### **The evolution of primary care systems**

Information processing started in scientific and hospital centres in the early 1950s. In most countries, GPs had uniform wishes regarding computer applications. In the sixties, GPs in the United Kingdom used limited minicomputer systems (in terms of storage and memory capacity)<sup>1,2</sup>. By the mid 1970s, some type of automated primary care patient record system<sup>3-7</sup> had been implemented at various sites in Europe. These developments could perform simple administrative tasks and some specific functions like screening for vaccinations, or were try-out structures concerning medical records. At that time, the first concept of a problem-oriented recording structure was presented. Problematic issues, like costs, data security, and undisciplined record keeping and data collection, stagnated progress in GP computing. The introduction of the personal computer in the 1980s and especially their relatively low costs promoted computing in general practice<sup>8-11</sup>. Methods for collecting, processing, analysing, and reporting data accurately were developed<sup>12</sup>. Today, a great variety of applications in primary care exists<sup>13-20</sup>. These applications comprise facilities for e.g. communication, community health care statistics, diagnostic decision support, and medical record keeping.

#### **Primary care patient information systems**

##### *Reasons for development*

Several reasons can be identified for the development of patient information systems. These are that GPs want:

- to work more effectively and efficiently,
- to achieve a higher quality of care, and
- to give disease management using software (this requires standardization in terms of e.g. coding and definitions of diseases).

##### *From PPRs to CPRs*

GPs capture medical data during daily activities (patient encounters and/or visits). Data registration is needed for the management of ongoing medical care. By registering patient encounter data, GPs can evaluate medical decisions retrospectively. These data,

namely, reflect a GP's decision-making process.

Initially, paper was used as a means for registering data. Nowadays, there is a trend to replace Paper-based Patient Records (PPRs) by Computer-based Patient Records (CPRs). PPRs, namely, have significant limitations in hospital, like unavailability, a huge physical volume over time, unreadability, difficulties regarding the transferability and the integration of data between health care providers and institutions, and the necessity of having to record the same data many times on different documents<sup>21</sup>. The same limitations, yet not as extensively, occur in primary care. During holidays, e.g., a GP's replacement can be faced with problems of unreadability or unavailability of PPRs. The volume of some patients' PPRs including letters, etc., can also be bulky in general practice. The necessity of having to record the same data on different documents (e.g. concerning tests performed at Laboratory Departments), furthermore, can be a problem to GPs. The increasing need for disease management (shared care) and the necessity of information management have proven the inadequacy of PPRs. CPRs may be the solution.

The CPR concept is described in literature<sup>22-24</sup>. CPRs are equipped to store large amounts of data from different sources. Gathering and entering medical data in an automated system can make the registration process more structured<sup>25</sup>. For communication and research purposes, standardization of medical data by using a standardized vocabulary is needed. Free-text processing, namely, causes a lot of interpretation problems to automated systems.

Data entry and data presentation may have different requirements. In a hospital environment, it was investigated how information should be organized and displayed to facilitate an optimal retrieval of information from CPRs<sup>26</sup>. It was concluded that the use of headings in providing contexts made the search for information faster.

Many initial difficulties of CPR keeping have been solved. A lot of time and costs in terms of personnel and resources, for instance, were needed to achieve and maintain accurate and reliable information. The quality of CPRs implemented in early systems, furthermore, was criticized in that these records were frequently poorly organized<sup>18,27-31</sup>. With key aspects of care missing and inaccurate diagnostic coding<sup>32,33</sup>, they could be inaccessible, incomplete, and inaccurate<sup>27,34,35</sup>. Errors could be introduced by clinicians, patients, or equipment<sup>36</sup>. A change in routine clinical practice, involving for instance the use of structured data entry sets<sup>37</sup>, was needed to solve many of the mentioned difficulties.

### *GP participation*

It has long been recognized that "computers could and should be used in medical record keeping"<sup>38</sup>. Record keeping is a required function in a GP information system. However, primary care patient information systems are useful and successful only if GPs are involved in the development process. One way to involve GPs is to explore their expectations and requirements regarding these systems<sup>39</sup>. A distinction is made between technical and functional system requirements: functional requirements concern a system's functions, whereas technical requirements refer to how functions are put into practice. According to the authors<sup>39</sup>, emphasis should be on technical wishes, for a system will not be used if it is not easy to use, regardless of its functional capabilities. Technical requirements mostly expressed by GPs are that a system should: improve their and the practices' efficiency; save time; reduce costs; reduce the workload; and have a good system-user interface. Functional requirements refer to practice management, patient management, and quality assurance. Practice management concerns all administrative processes in a GP practice. Patient management refers to data collection and to all processes which are beneficial to patients (it should e.g. be possible to check on drug interactions, to register medical histories, and to monitor patients). Quality assurance includes refresher courses as well as audit and planning.

### *SOAP registration and episodical recording*

In order to facilitate structured data capture, data storage, and data retrieval, methods were advised by Colleges of General Practitioners and accepted by user groups. Two of these methods are problem-oriented recording, i.e. the categorization of encounter data using the method of SOAP registration, and episodical recording<sup>40-44</sup>, i.e. data concerning a disease or a problem of a patient are gathered and grouped together during a disease period. Every time a disease or problem recurs, a new period is initiated which, via the episodical organization, is linked to the previous period. Thus, an entire overview of episodes of the disease or problem is provided over time. SOAP is the acronym for Subjective findings (symptoms stated by a patient), Objective findings (signs found in medical examination), Assessment (diagnosis), and Plan (therapy), and is a method for linking a patient's clinical data to problems<sup>22,23,45-47</sup>. Additionally coding instruments were developed, like the International Classification of Primary Care (ICPC) and the International Classification of Health Problems in Primary Care (ICHPPC)-II-Defined which are both accepted by the World Health Organization.

*Effects of primary care patient information systems*

Developments in software and hardware technology and an increased computer-mindedness of GPs have led to a growth in automated support and a diversification of computer applications. Computers in general practice can have positive effects, like:

- a better data capture, management of information, and exchange of data between disciplines,
- an increased effectiveness (faster and better care),
- an increased rationality in diagnostic reasoning and in the use of resources,
- a better insight in the level of patient care and in health care epidemiology,
- an improved cost awareness of health care, and
- a better level of education.

The ability of computers to use data in multiple ways, furthermore, enables quality control<sup>48</sup>. Information systems meet the increasing demands of health care providers and patients, and are the principal instruments for ensuring coordinated, continuous, and comprehensive care<sup>49</sup>.

*Reference guides in The Netherlands*

The reference guide<sup>50</sup> for GP information systems published by the Dutch College of General Practitioners in 1985 contained a description of administrative data for the patient registration module, which is one of the modules in a GP information system. This module enables functions like: 1) billing, 2) the registration of patient names, addresses, and other patient-related data, and 3) the registration of background information (social class, religion, living conditions, vaccinations, etc.). The medical module, which consists of a patient encounter module, a problem list module, and a medication module, was only mentioned as being relevant to future GP information systems and described very briefly. In the 1995 reference guide<sup>51</sup>, the medical module is well described and extended with modules for communication, archiving diagnostic data of e.g. laboratory tests, disease prevention, medication control, episodic recording, and the presentation of guidelines in text format.

*Present situation in The Netherlands*

According to the Dutch College of General Practitioners, about 90% of GPs in The Netherlands today use computers. Whereas only 25% of these GPs apply ICPC coding in the medical record, 80% of them employ the drug prescription and control module, and 50% use the SOAP approach. It is generally expected that eventually CPRs



will replace PPRs. Thus, progress is being made towards the right direction. Adequately structured medical records are believed to improve individual patient care as well as the health care system in general. It is therefore important that an effort is made to obtain such adequately structured and accurate medical records and to make them easy to handle and maintain and readily available. However, arrangements should also be made to warrant the confidentiality of CPRs<sup>52</sup>.

### **Decision Support for General Practitioners**

#### **System evolution**

In the fifties and the early sixties, information systems were used mainly as tools for processing data. These systems were called Transaction Processing Systems (TPSs)<sup>53,54</sup>. In the late sixties and the seventies, the function of computer information systems gradually changed: the importance of large data bases (together with data base management software) and managerial information became apparent. Management Information Systems (MISs) - computer information systems which process and provide managerial information<sup>53-55</sup> - were built on top of TPSs. In the eighties, managers indicated that they were not satisfied with information systems which merely presented information in a very efficient way. They became increasingly aware of the need to be supported in their decision making by computer information systems. Hence, DSSs and Expert Systems (ESs)<sup>53,54</sup> were introduced, having the following distinguishing functions:

- DSSs actively interact with their users. These users can provide additional information, and the DSSs eventually present decisional alternatives from which the users can choose<sup>56-58</sup>.
- ESs do not need active user participation during the decision process, apart from being provided with initial data by their users, and only present the decisional alternatives which have the highest statistical probability or are most relevant<sup>58</sup>.

#### **The development and the implementation of DSSs**

Much has been written about the development and the implementation of DSSs<sup>56-58</sup>. Especially in the nineties, these systems were increasingly used.

#### ***DSS definition***

Many researchers have defined a DSS<sup>56-62</sup>. Definitions differ on details only and not on the principle issue. The most common description is that a DSS is an information system which interactively supports decision making<sup>56,57</sup>. In order to provide adequate

decision support, a DSS minimally should contain the following components<sup>56,57</sup>: a data base (in which data are stored which need to be processed), a model base (in which knowledge is stored in the form of models), and dialogue and management software (which respectively provides user/system interaction and keeps track of the entire decision process). The decision support given includes the presentation of additional information or decisional alternatives. A DSS arrives at these alternatives on the basis of data stored in the data base and knowledge stored in the model base. A user evaluates and validates the presented alternatives, and eventually chooses one. Users, therefore, are essential in automated decision support.

#### *The decision-making process and modelling*

The rational decision-making process was first analysed by Simon<sup>55</sup>. Three decision-making phases are distinguished: intelligence, design, and choice. In the intelligence phase, the various aspects of a problem in the field of application are analysed. Alternative solutions to the problem are described in the design phase. In the choice phase, the best possible decisional alternative is chosen. Simon's model should be extended with a fourth phase: implementation<sup>54</sup>. In the implementation phase, a decision maker puts the decision to the test and evaluates its effects. If the decision does not live up to the expectations, it should be reviewed and a new rational decision-making process should be initiated.

The process of creating a representation of the real world in the form of models is called modelling<sup>63-65</sup>. Modelling includes knowledge representation and knowledge presentation. Knowledge representation is defined as the way in which knowledge is organized in models. Knowledge can be represented through logical expressions (rules/frames) like Medical Logic Modules (MLMs), object-oriented representations, and probabilistic knowledge (Bayes' theorem)<sup>66-71</sup>. Knowledge presentation is defined as the process of presenting models and other data stored in an information system to a user. Knowledge presentation is done with the help of windows, icons, mice, light pens, voice synthesis, etc.

#### *Types of decision support*

Several types of decision support can be identified. In the most basic differentiation, active decision support is distinguished from passive decision support<sup>56,57</sup>. Whereas passive decision support is confined to the presentation of information, active decision support includes reasoning. Active decision support, in turn, can be subdivided

into goal-oriented reasoning and data-oriented reasoning<sup>53,54</sup>. Although in some fields of application DSSs can be used which apply either goal-oriented or data-oriented reasoning most application areas require a combined strategy of these two reasoning types<sup>53,54</sup>. Consequently, there is a growing need for DSSs capable of both goal-oriented and data-oriented reasoning.

### *DSS development*

In order to obtain the functional specifications of a DSS, methods and techniques<sup>5</sup> can be used like prototyping, the top-down and/or bottom-up technique, and iterative design. Prototyping involves interviewing future users to assess the user environment determine the type and quality of decisions made and their impact on the field of application by evaluating each prototype, and obtain system requirements. System requirements can also be obtained with the top-down and/or bottom-up technique, in which the developers' and the users' scopes are changed several times from bottom up to top down or vice versa during the analysis of a problem. Iterative design implies that several prototypes of a system are consecutively implemented at a user site and that each one of the prototypes is updated and extended with hardware and/or software on the basis of users' remarks.

Simon's decision-making model can be the basis for modelling decisions with respect to the development process, the implementation process, and the problem field. Potential users should be well instructed and supported prior to, during, and after the implementation of DSSs at user sites. It should be made clear that users have nothing to fear but everything to gain by DSSs.

### *DSSs in hospital and in primary care*

DSSs can be used in a wide range of application areas<sup>72-83</sup>. Examples of systems implemented in hospital are listed in table 1.

Prior to 1990, no integrated primary care DSS existed. Integration is needed because otherwise data have to be entered more than once. GPs also favoured an integration of decision support facilities with a GP information system. Since the nineties, consequently, several primary care DSSs have been developed. The Hypercritic system, e.g., used the downloaded patient data base of the GP information system ELIAS to provide support regarding the prescription of drugs<sup>84</sup>. Use was made of therapy protocols for hypertension. AAH Meditel's Clinical System<sup>85</sup> and VAMP medical<sup>86</sup>, both designed in the United Kingdom, are examples of primary care DSSs in which diagnostic decision

support facilities are integrated with a GP information system. The same is the case with SWEDESTAR which provides drug prescription support as well as diagnostic support regarding chronic diseases (hypertension, asthma, diabetes, etc.)<sup>87</sup>. It can be concluded that only a small number of integrated DSSs are operational in the primary care environment.

**Table 1. Some hospital DSSs and their application areas**

Name	Diagnostic or Therapeutic	Application areas
ABEL	D	Acid-base electrolyte imbalances
AI/RHEUM	D	Connective tissue disorders in clinical rheumatology
ARAMIS	D	Rheumatic diseases
CASNET/GLAUCOMA	D/T	Disease states related to glaucoma
DIABETEX	T	Type-I diabetes
EXAMINER	D	Examination of the diagnostic competence of internal-medicine consultants
HELP	D/T	Critiquing system for diagnostic and therapeutic advice
INTERNIST-I/ CADUCEUS	D	General diagnostics on the basis of a patient's history, symptoms, and laboratory results
IRIS	D	Ophthalmology
KADIS	T	Type-I diabetes
MYCIN	D/T	Bacteraemia, septicaemia, meningitis, and cystitis
NEOMYCIN	D/T	Meningitis and similar diseases
NEUROLOGIST-I	D	Neurological disorders and the localization of lesions in the central nervous system
PROTOVIEW	D/T	Passive support regarding bone fractures in the Emergency Department and obstructive jaundice
QMR	D	Internal medicine
REGENSTRIEF	D/T	Reminder system using medical records

### The evaluation of DSSs

The evaluation of a DSS is considered essential during its entire design process. The range of evaluation issues has widened to a multidisciplinary perspective. DSSs are evaluated from ethical, legal, economic, sociological, psychological, medical, and

intellectual points of view<sup>88-92</sup>. There are many evaluation methodologies. None of the methodologies, however, has been commonly accepted. Most of them discern three phases<sup>88</sup>: the definition phase, the laboratory phase, and the field test phase. The definition phase is in fact an iterative build-test-refine cycle. In the laboratory phase users and system developers evaluate a DSS prototype from their perspectives in a laboratory setting. In the field test phase, the system's effectiveness in the field is studied. A DSS should be evaluated on the basis of the following criteria: the validity of the knowledge stored in the model base, the accuracy of the results given by an individual DSS module and the entire system (Does a DSS module produce the results intended when used in isolation, and do all DSS modules in a system produce the intended result when working together?), the usefulness and the userfriendliness of the system, and the effectiveness of the system in improving the application process when it is used in its intended environment<sup>92</sup>. To evaluate the quality of a DSS, a golden standard is needed. The stipulation of such a golden standard often appears to be a difficult task. How difficult this can be is illustrated by a study concerning the quality of drug prescription for epilepsy patients. In this study, 5 neurologists were unable to reach consensus using a Delphi method<sup>93</sup>.

Over the years, there have been successful introductions of DSSs, but also failures<sup>89-91</sup>. A DSS is considered successful if it meets a practical need; if it helps alleviate the problem of information overload; if it has an acceptable target group of users; if it saves time; if users do not feel threatened by the system as it increases their productivity, lets them in charge of things, or avoids possible legal problems; if it makes the knowledge domain less fuzzy; and if it has a positive impact on the field of application<sup>89</sup>. Reasons for failure of a DSS have been that there was a large mismatch between the system's functions and the real problems users faced; that the role of the provided decision support was unclear to the majority of potential users (a communication gap existed between the users and the system designers); that no coherent development philosophy existed; and that organizational issues were disregarded<sup>89</sup>. Many of these problems can be solved by establishing and clarifying the need for a DSS through prototyping; by using a coherent methodology for modelling a problem; and by applying sound methodologies for the iterative development, implementation, and evaluation of a DSS<sup>89</sup>.

### **The impact of DSSs on medicine**

Since their introduction, medical DSSs are to a certain extent being used in all

kinds of medical application areas. Much has been written about their use in and possible effects on health care<sup>76,94</sup>. One of their very important benefits is an improvement of medical effectiveness<sup>94-96</sup>.

Medical effectiveness had already been improved greatly by the introduction of information systems, in which users can store medical and patient data fast and efficiently. Different ways for assuring the quality of a data base and of data stored in it<sup>97,98</sup> proved to be effective and therefore enhanced the usefulness of data bases. Much, however, depends on user attitude and involvement. The introduction of software in organizations like general practices or health centres will lead to a change in work activities as well as responsibilities. The entire staff involved needs to be convinced of the attainability and the necessity of the changes. The discipline of the staff can be important for the well-functioning of a decision support technique<sup>99</sup>. The staff, namely, needs to provide the necessary input and for instance can direct a DSS<sup>58-60</sup> in a certain direction in case of a conflict situation.

DSSs can offer advice and warnings in decision-making processes regarding diagnosis and therapy determination<sup>76,95,96</sup> and in decision-making instances like the determination of the necessity of laboratory test requests of medical professionals<sup>100</sup>. In the latter application, a DSS can also have a considerable impact on health care since medical professionals request many unnecessary laboratory tests which is rather time-consuming and costly.

As medical DSSs become more widespread, their evaluation regarding diagnostic accuracy and usefulness becomes more eminent. In a study<sup>101</sup>, four internal medicine DSSs were compared with regard to their accuracy. It appeared that none of the four diagnostic DSSs was consistently better or worse than the others. All produced moderately long lists of potential diagnoses. The lists included many diagnoses which a knowledgeable internist would regard as not particularly helpful in explaining a case. Each program on the other hand also suggested some diagnoses, though not highly likely ones, of which the experts later on agreed that they were worthwhile including in the differential diagnosis. The study raised the concern that important diagnostic considerations may be that obscured by other diagnoses that the value of the programs is significantly decreased, or that the programs in inexperienced hands will lead to excessive or costly interventions.

### Medical Knowledge of Thyroid Disorders

#### The test ordering by GPs

The expenditure on laboratory tests in primary care is constantly increasing - as it has been doing each year in the past decades - due to a rise in the number of diagnostic tests requested by GPs. The rise is a result of a combination of aspects like an ageing population and therefore a need for more requests, an increased awareness of patients to make use of the health care system, and advances in technology. A defensive attitude of GPs in the diagnostic process is another aspect which leads to an increase in test requests. GPs do not always know the indications for requesting tests; there sometimes is a lack of knowledge concerning principles of medical decision making. When requesting a test, a GP should take into account the following decision-making principles:

- a sufficiently high pretest probability of the disease,
- a choice of tests on the basis of their characteristics (sensitivity, specificity, predictive values, likelihood ratios), possible complications, and costs involved, and
- whether or not there are useful consequences of initiating a policy.

Sometimes GPs request tests just to be on the safe side. There is therefore a need for guidelines concerning test adequacy. In order to increase the rationality of the request behaviour, several methods can be applied. One of them is giving feedback to GPs on their test-ordering behaviour. The effects of this kind of feedback were investigated in several studies<sup>102-111</sup>. Feedback was found to result in a reduction in test requests and, thus, in cost savings. However, regular feedback was needed to ensure the continuity of these effects. Moreover, providing feedback is time-consuming and costly. It was ever doubted if feedback was economically worthwhile<sup>112</sup>; costs for providing feedback should be carefully weighed against expenditures saved and against costs involved in a late detection of patients' diseases.

The Diagnostic Coordinating Centre (DCC) of the University Hospital Maastricht investigates the adequacy of diagnostic test requests of GPs. About 87 'Heuvelland' (Maastricht region) network GPs, working in a catchment area of approximately 191,000 persons, are given individual feedback regarding their request behaviour in a randomly selected month within a six-month period. In order to enable an assessment of the adequacy of their test requests, the GPs provide the DCC with relevant data by completing a paper-based request form for each patient for whom tests are requested. Questions on this form concern: general data on the patient and the requesting GP, the reason for a request, possible hypotheses, signs/symptoms, physical examination results,

medication, the past history, and tests to be performed. The last category comprises chemical blood tests, haematological tests, virological and serological tests, bacteriological tests, faecal and urine tests, electrocardiograms, and diagnostic imaging (radiology, echography, endoscopy). The feedback on the basis of an analysis of the GPs' request forms is given by a respected expert peer with a background in internal medicine. The feedback comprises an overview in terms of the number of tests requested by a GP, and a discussion of the rationality of the requests. The rationality is assessed using regional and national guidelines as well as the principles of medical decision-making. Regional guidelines were designed and agreed upon by the participating GPs, whereas national guidelines were stipulated by the Dutch College of General Practitioners. The feedback deals with inappropriate requests of tests and contains recommendations for a more rational diagnostic request behaviour. Reference is made to the guidelines and the data provided on the request forms.

The DCC's regular feedback given since 1985 has positive effects. The number of requests of tests frequently addressed in several feedbacks as being inappropriate given the clinical pictures of patients decreased significantly<sup>113</sup>. The regularly provided reminders also led to a reduction in requests of other tests. The feedback, furthermore, is found to be economically worthwhile. From 1985 until 1991, a cumulative cost saving of \$ 1,369,621 was obtained in the Maastricht region because of test reductions. Cost savings came to about \$ 369,000 yearly in the past five years. This resulted in a cumulative saving of about \$ 3,214,000 in 1996. It can be concluded that the feedback not only results in a reduction in the number of requests<sup>114</sup>, a consequent cost reduction<sup>115</sup>, and an improvement of the quality of requests<sup>107</sup>, but also in a change in GP attitude towards a more cost-effective 'style of practice'<sup>116</sup>.

### **Past and present tools for diagnosing thyroid disorders**

In the Maastricht region, GPs request thyroid tests on the basis of work procedures, which are presented in Appendix I. In these procedures, thyroid disorders are described and advice is given concerning the appropriate request of thyroid function tests.

The number of requests of a Thyroid-Stimulating Hormone (TSH) test, a sensitive and specific test for the diagnosis of thyroid disease and one of the tests on the paper-based request form, increased in recent years. The majority of the requests seemed to be irrational. Often, namely, GPs based their TSH test requests on only one or a combination of a few atypical sign(s)/symptom(s). If a pretest probability of a thyroid disorder is determined on the basis of these data, it will be low.



To answer questions like 'Is subclinical hyperthyroidism a disease?' and 'Does it need to be treated immediately?' more should be known about the natural course of thyroid disorders. That only little is known was assessed by means of a literature study and by consulting medical experts. In literature, only two studies<sup>117,118</sup> were found concerning the evolution of subclinical hyperthyroidism or hypothyroidism into the clinical types. In these studies, the applied patient groups were small, either involving 4 elderly (53-98 years) or 18 young (5-19 years) patients.

Thyroid disorders are fairly common diseases. Incidence rates assessed in a Dutch national study<sup>119</sup> of diseases in general practices suggest that the total population of 191,000 persons of the 87 GPs annually comprises 153 (0.8 per 1,000) new clinical hyperthyroid and 76 (0.4 per 1,000) new clinical hypothyroid patients. Regretfully, the incidence rates provided by the Registration Network General Practices of the university Department of Family Medicine could not be used due to contradictory outcomes. These outcomes indicated that the incidence rate of hypothyroid patients was twice as high as that of hyperthyroid patients, which is not in agreement with literature and textbook incidence rates. A British incidence rate indicates that a regular GP in the United Kingdom detects 7 new thyroid patients a year in a general practice population of on average 2,500 persons<sup>120</sup>. On this basis, 535 new thyroid patients should be expected each year. It is, however, questionable whether the British rate is applicable to the population in the Maastricht region.

Unnecessary test requests<sup>121,122</sup> can be prevented by using the principles of medical decision making within which the determination of a patient's pretest probability of disease and the discriminative power of diagnostic tests, e.g., are important factors. A solution for GPs who have difficulty in rationally ordering TSH tests, therefore, can be to provide GPs with decision-making instruments (syn.: indices, scoring instruments, models, scoring questionnaires, prediction instruments) for thyroid disorders. These instruments should predict the probability of a positive test (and consequently of the presence of a disease) using signs/symptoms present in a patient. A probability threshold should be applied to decide whether the test needs to be performed. The implementation of the prediction instruments in a DSS provides GPs with direct feedback<sup>123,124</sup> which can make the DCC's time-consuming retrospective manual feedback superfluous.

A literature review showed that no thyroid disorder prediction instruments exist for the specific use by GPs; existing instruments were all designed for hospital application and concern the diagnosis of clinical hyperthyroidism or clinical hypothyroidism on the basis of signs/symptoms which are each awarded a fixed score. In the past, hospitals

indices were frequently used in clinical practice all over the world<sup>125-129</sup>. Prior to this use of indices, the clinical picture (medical history and physical examination data) in combination with the Basal Metabolic Rate (BMR) were used to diagnose a thyroid disorder, more or less as a result of the lack of hormone and immunologic tests.

A historical overview shows that over the years the BMR (until 1974), the Free Thyroxine Index (FT<sub>4</sub>I; 1974-1978), the Free Thyroxine (FT<sub>4</sub>) laboratory test (1978-present), the 1<sup>st</sup> generation TSH Radio-ImmunoAssay (TSH RIA) test (1983-1985), and the 2<sup>nd</sup> generation TSH ImmunoRadioMetric Assay (TSH IRMA) test (1985-1994) were successively applied in thyroid diagnostics. Nowadays (1994-present), the 3<sup>rd</sup> generation TSH ImmunoFluoroMetric Assay (TSH IFMA) test and TSH ImmunoChemiluminescent Metric Assay (TSH ICMA) test as well as other nonradioactive labels are used.

**Table 2. Indices for clinical hyperthyroidism and hypothyroidism**

Hyperthyroidism	Patient numbers learning/test set	Method	Golden standard	Sensitivity learning/test set	Specificity
Crooks/Murray/ Wayne	182 (?)/121 (50)	Lin. Discr. Anal.	BMR	88 %-99 %/ 86 %	83 %-97 %
Wayne	-	Lin. Discr. Anal.	BMR	-	-
Börner	433 (68)	Lin. Discr. Anal.	BMR	98.5 %	94.2 %
Gurney	136 (44)	Lin. Discr. Anal.	BMR	64 %	89 %
Fragu	410 (112)	Bayes' theorem	BMR	74 %	30 %
Von Laubinger	313 (129)	Lin. Discr. Anal.	BMR	90.7 %	94.6 %
Möpert/Vollmar	548 (185)	Lin. Discr. Anal.	BMR	92.7 %	71.7 %
Simonin	249(120)/3596(?)	Lin. Discr. Anal.	T <sub>3</sub> /T <sub>4</sub>	92.5 %	98 %
Hypothyroidism	Patient numbers learning/test set	Method	Golden standard	Sensitivity learning/test set	Specificity
Wayne	150 (100)	Lin. Discr. Anal.	BMR	98 %	100 %
Billewicz	96 (31)/31 (25)	Multiple Regres.	BMR	100 %	41 %
Seshadi	52 (12)	Lin. Discr. Anal.	BMR	92 %	45 %
Murray	242 (127)	Lin. Discr. Anal.	BMR	88 %	93 %

### A review of indices

In literature, twelve hospital indices for the diagnosis of clinical hyperthyroidism or clinical hypothyroidism were found. They contain items which are awarded points

depending on their presence or absence in a patient. If the total of points awarded is higher than a certain threshold, the test is requested. In table 2, the following aspects (if mentioned in the consulted literature) are described for each index: the number of patients

Table 3. Diagnostic indices for clinical hyperthyroidism

	Crooks/Murray/Wayne		Wayne		Börner		Gurney	
	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Dyspnoea on exertion	+1		+1					
Heart palpitations	+2		+2		+1			
Tiredness	+2		+2					
Preference for heat		-5		-5				
Preference for cold	+5		+5					
Excessive sweating	+3		+3					
Nervousness	+2		+2					
Appetite increased	+3		+3				+5	
Appetite decreased		-3		-3				
Weight increased		-3		-3				
Weight decreased	+3		+3					
Palpable thyroid	+3	-3	+3	-3				
Thyroid murmur	+2	-2	+2	-2			+18	
Exophthalmos	+2		+2		+1		+9	
Lid retraction	+2		+2				+2	
Lid lag	+1		+1					
Warm hands	+2	-2	+2	-2	+1			
Moist hands	+1	-1	+1	-1				
Auricular fibrillation	+4		+4					
Pulse rate < 80/min		-3		-3				
Pulse rate 80-90/min							+8	
Pulse rate > 90/min	+3		+3				+16	
Hyperkinetic movements			+4	-2			+4	
Fine finger tremor			+1		+1		+7	
Pulse rate ≥ 100/min					+1			
(Systolic-diastolic blood pressure) ≥ 60 mmHg					+1			
Anxiety							-3	
Frequent checking							-3	
Psycholog. precipitant							-5	
Palpable thyroid							+3	
Age 15-24								
25-34							+4	
35-44							+8	
45-54							+12	
55 and over							+16	

in the learning and test sets - the number of thyroid patients in a corresponding set is presented between parentheses; a question mark indicates that this number could not be inferred -, the analysis method(s) used to create the indices, the laboratory test performed (used as the golden standard), and the sensitivity and specificity of the indices. As can be

seen from this table often no separate test set was used to determine the sensitivity and the specificity. The indices are described in tables 3 through 5. A space in these tables indicates that an item is awarded zero points.

#### *Diagnostic indices for clinical hyperthyroidism*

Crooks, Murray, and Wayne (1958-1959)<sup>130,131</sup> developed the index described in table 3. In this table, also Wayne's index (1960)<sup>132,133</sup>, the index developed by Börner and colleagues (1968)<sup>134</sup>, and the Newcastle (Thyrotoxicosis) Index created by Gurney and colleagues (1970)<sup>135</sup> are presented.

The indices designed by Von Laubinger (1986)<sup>136</sup>, Fragu and colleagues (1979)<sup>137</sup>, and Simonin and colleagues (1982)<sup>138,139</sup> are described in table 4. Von Laubinger favoured Crooks' comprehensive index over Börner's brief one because it investigates the entire clinical picture of clinical hyperthyroidism. Nevertheless, Von Laubinger was not satisfied with that index and created one to be used in his own environment. Möpert and Vollmar (1970)<sup>140</sup> developed their index to avoid unnecessary tests in their Nuclear Medicine Department.

**Table 4. Diagnostic indices for clinical hyperthyroidism (continued)**

	Von Laubinger		Fragu		Möpert/Vollmar		Simonin	
	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Dyspnoea on exertion							-0.4	
Heart palpitations	+1				+1		7.2	
Tiredness	+2							
Preference for heat	-5		+1				0.6	
Preference for cold	+4						-14.6	
Excessive sweating	+2						0.8	
Nervousness			+1				-0.3	
Appetite increased	+3						0.7	
Appetite decreased	-3							
Appetite changed			+1					
Weight increased	-3						-2.4	
Weight decreased	+3				+1			
Weight loss > 5 kg			+1					
Palpable thyroid	+2	-1						
Thyroid murmur	+3	-1			+1		1.7	
Exophthalmos	+2				+1			
Other eye problems	+2							
Auricular fibrillation	+4							
Skin warm	+3	-2	+1					
Increased frequency of defaecation	+3							
Obstipation	-3						-5.8	
Hair loss	+1				+1			
Hyperkinetic movements	+3						1.0	

## Chapter 2

(continued)	Von Laubinger		Fragu		Möpert/Vollmar		Simonin	
	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Fine finger tremor	+1						0.6	
Pulse rate < 80/min	-3						-5.2	
Pulse rate 80-90/min								
Pulse rate > 90/min	+3		+1				3.7	
Pulse rate > 100/min					+1			
Lid retraction			+1					
Thirst increased			+1				2.0	
Fine finger tremor			+1		+1			
Diarrhoea					+1		3.6	
(Systolic - diastolic blood pressure) $\geq$ 60 mmHg					+1			
Noisy abdomen					+1			
Dry/moist skin					+1			
Warm hands							0.7	
Moist hands							0.2	
Warm and moist hands							0.3	
Cold hands							-5.5	
Cold and moist hands							4.5	
Dry skin							-4.5	
Face swollen							-2.0	
Insomnia							-1.0	

### Diagnostic indices for clinical hypothyroidism

The diagnostic indices for clinical hypothyroidism designed by Wayne (1960)<sup>141</sup>, Billewicz and colleagues (1969)<sup>142</sup>, Seshadi and colleagues (1989)<sup>143</sup>, and Murray (1964)<sup>126</sup> are described in table 5.

**Table 5.** Diagnostic indices for clinical hypothyroidism

	Wayne	Billewicz		Seshadi		Murray	
	Present	Present	Absent	Present	Absent	Present	Absent
Dry skin	+3	+3	-6	+3	-6	+3	-1
Constipation	+2	+2	-1	+2	-1	+3	-1
Hoarseness	+3	+5	-6	+5	-6	+2	-1
Tiredness	+1		+2		+2	+1	
Lethargy	+3					+3	-3
Paresthesia	+2	+5	-4	+5	-4	+2	-1
Deafness	+2	+2		+2		+1	
Preference for heat	+1	+4	-5	+4	-5	+1	
Preference for cold	-3					-3	
Decreased sweating	+2	+6	-2	+6	-2	+2	
Increased sweating						-3	
Hair loss	+2					+2	
Dry hair				-2	+2	+1	
Appetite increased	-3					-2	
Appetite decreased	+2					+1	
Weight increased	+1	+1	-1	+1	-1	+2	

(continued)	Wayne	Billewicz		Seshadi		Murray	
	Present	Present	Absent	Present	Absent	Present	Absent
Weight decreased	-3					-2	
Slow cerebration	+2	-3	+1	-3	+1	+2	-1
Slow movements	+2	+11	-3	+11	-4	+2	-2
Skin coarse	+2	+7	-7	+7	-7	+1	
Skin cold	+2	+3	-2	+3	-2	+2	-2
Skin pale	+2					+1	
Malar flush	+2						
Periorbital puffiness	+2	+4	-6	+4	-6	+2	-2
Puffiness							
supraclavicular	+2					+2	-1
Pulse rate < 61/min						+3	
Pulse rate 60-69/min	+1	+4				+1	
Pulse rate 70-79/min		+4					
Pulse rate > 79/min	-2	-4				-2	
Pulse rate $\geq$ 75/min				+4	-4		
Ankle jerk reduced		+15	-6	+15	-6		

### Laboratory tests

In the past, several circumstances considerably influenced the quality of the laboratory determination of the thyroid function. A patient's medical history of comorbidity, drugs, living conditions, and external factors are known to influence laboratory tests<sup>144</sup>. Many tests in use prior to the TSH IRMA test were difficult to interpret<sup>145-147</sup> and at least 3 days were needed to analyse an outcome. With the introduction of the TSH IRMA test, it became possible to differentiate between hyperthyroidism, hypothyroidism, and euthyroidism<sup>148</sup> in about 1 day.

Sensitivities and specificities of the earlier tests - the BMR, the FT<sub>4</sub>I, the FT<sub>4</sub> test, the TSH RIA test, and the TSH IRMA test - were not as high as those of the TSH IFMA test, which is in use nowadays and constitutes especially a technological improvement in that lower concentrations can be measured. Since the TSH IFMA test came in use after the studies described in chapters 8 and 9 already were initiated, the TSH IRMA test was applied in these studies. Sensitivities, specificities, positive Predictive Values (PVs+), and negative Predictive Values (PVs-) of the different laboratory tests were determined in various studies and are presented in table 6 along with the numbers of patients used in the studies. In most of these studies, no distinction was made between hyperthyroidism and hypothyroidism. Table 6 shows that most studies were carried out on hospital patient groups. Two outpatient studies were performed. The research by Klee was based on 87 untreated hyperthyroid patients with Graves' disease (18 males, 69 females, age: 15-89 years, median: 47 years) and of 364 euthyroid patients (165 males, 199 females, age: 10-92 years, median: 56 years). The high PV+ can be explained by the fact that 19.3% of

the patients for whom TSH tests were requested in hospital had thyroid problems. In the Nuuttila study, a patient group of 354 patients was used. Of these, 130 were outpatients (25 males, 105 females, age: 21-85 years, mean age: 51 years) and 224 primary care patients (53 males, 171 females, age: 5-99 years, mean age: 52 years). Eighteen patients had a past (more than 6 years ago) history of thyroid resection. Patients obtaining a T<sub>4</sub> replacement therapy were excluded. The low PVs+ could be expected for the primary care environment due the large number of euthyroid patients, the small number of thyroid disorder patients, and the applied laboratory test which has a specificity less than 100%. The golden standard used for corroborating the laboratory tests was a combination of clinical grounds based on signs/symptoms presented by patients, previously used laboratory tests (Total T<sub>4</sub>, Total Triiodothyronine (T<sub>3</sub>), FT<sub>4</sub>, FT<sub>3</sub>, the Thyroid-Hormone Binding Ratio, the Thyrotrophin-Releasing Hormone (TRH), scintigrams), and the reproducibility of one or several test outcomes. The sensitivities and the specificities of the various tests listed in table 6 constantly improved over time.

**Table 6.      Various laboratory tests and their values**

Source	Laboratory tests	Population	Patients euth hr/ho		Sensitivity %	Specificity %	PV+ %	PV- %
Caldwell <sup>145</sup>	TSH IRMA	hospital	97	188	99	100	100	99
Klee <sup>146</sup>	TSH IRMA	hospital	142	87	100	99	99	100
	TSH IRMA	outpatient care	364	87	100	98	92	100
Henneman <sup>148</sup>	TSH IRMA	hospital	249	255	96	94	94	96
	FT <sub>4</sub>				98	87	89	87
	FT <sub>4</sub> I				89	96	96	90
Mudde <sup>149</sup>	TSH IRMA	hospital	14	22	94	100	100	95
	TSH RIA				65	95	92	75
	FT <sub>4</sub>				82	95	94	86
	FT <sub>4</sub> I				53	95	89	69
Nuuttila <sup>150</sup>	TSH IRMA	outpatient care	111	19	100	84	51	100
	FT <sub>4</sub>				84	95	73	97
	T <sub>4</sub>				89	87	55	98
	TSH IRMA	primary care	218	6	100	85	16	100
	FT <sub>4</sub>				50	96	25	99
	T <sub>4</sub>				33	91	9	98
Hay <sup>151</sup>	TSH IRMA	hospital	122	17	94	89	55	99
Ericsson <sup>152</sup>	TSH RIA	hospital	421	57	96	95	73	94
	FT <sub>4</sub> I				61	99	90	95
	T <sub>4</sub>				81	88	47	97
	T <sub>3</sub>				77	97	77	97

Table 7 presents the sensitivities, the specificities, and the PVs for clinical hyperthyroidism and clinical hypothyroidism as published by Coli<sup>153</sup>. In this study carried out in a hospital, 529 euthyroid, 127 hypothyroid, and 205 hyperthyroid patients participated. Results presented refer to the use of a single TSH test or a combination of the TSH and the FT<sub>4</sub> test. The TSH test proves to be a good tool for distinguishing euthyroidism from thyroid disorders.

**Table 7. Sensitivities, specificities, and PVs of the TSH IRMA test and/or the FT<sub>4</sub> test regarding hyperthyroidism and hypothyroidism**

Category	Sensitivity	Specificity	PV+	PV-
Hypothyroidism TSH IRMA	97	99	96	99
Hyperthyroidism TSH IRMA	99	99	97	99
Hypothyroidism FT <sub>4</sub>	86	99	92	98
Hyperthyroidism FT <sub>4</sub>	88	97	89	96
Hypothyroidism TSH and FT <sub>4</sub>	84	100	100	97
Hyperthyroidism TSH and FT <sub>4</sub>	87	100	100	96

## References

1. Dinwoodie HP. An elementary use of a computer for morbidity recording in general practice. *Health Bull Edinb* 1969;27(3):6-14.
2. Dinwoodie HP. Simple computer facilities in general practice - A study of the problems involved. *J Roy Coll Gen Pract* 1970;19(94):269-281.
3. Froom J. An integrated medical record and data system for primary care - Part 6: A decade of problem-oriented medical records - A reassessment. *J Fam Pract* 1977;5(4):627-630.
4. Froom J, Kirkwood R, Culpepper L, Boisseau V. An integrated medical record and data system for primary care - Part 7: The encounter form - Problems and prospects for a universal type. *J Fam Pract* 1977;5(5):845-849.
5. Parker A, Newell KW, Torfs M, Israel E. Appropriate tools for health care: Developing a technology for primary health care and rural development. *WHO Chron* 1977;31(4):131-137.
6. Grummitt A. Real-time record management in general practice. *Int J Biomed Comput* 1977;8(2):131-150.
7. Bain DJ. The results of developmental screening in general practice. *Health Bull Edinb* 1974;32(5):189-193.
8. Hosia P. Finstar - A comprehensive information system for primary care. *Scand J Prim Health Care* 1984;2(4):163-166.
9. Francis J, Roche M, Mant D, Jones L, Fullard E. Would primary health care workers give appropriate dietary advice after cholesterol screening? *Br Med J* 1989;298(6688):1620-1622.
10. Roos LL, Nicol JP, Cageorge SM. Using administrative data for longitudinal research: Comparisons with primary data collection. *J Chronic Dis* 1987;40(1):41-49.
11. Sigurdsson G, Einarsson I, Josafatsson JI, Magnusson G, Olafsson O, Sigvaldason H, Thorarinnsson S, Tulinius H. A medical record and information system for primary health care in Iceland - The



- Egilstadir Project. *Scand J Prim Health Care* 1984;2(4):159-161.
12. Ontario Ministry of Health - Technology Subcommittee of the Working Group on Critical Care. Guidelines for medical technology in critical care. *J Can Med Assoc* 1991;144(12):1613-1615.
13. Sullivan F, Mitchell E. Has general practitioner computing made a difference to patient care? - A systematic review of published reports. *Br Med J* 1995;311(7009):848-852.
14. Feder G, Griffiths C, Highton C, Eldridge S, Spence M, Southgate L. Do clinical guidelines introduced with practice based education improve care of asthmatic and diabetic patients? - A randomised controlled trial in general practices in East London. *Br Med J* 1995;311(7018):1473-1478.
15. Enlander D. *Computers in laboratory medicine*. New York: Academic Press, 1975.
16. Yamauchi K, Fukatsu T. A decision support system for diagnostic consultation in laboratory tests. In: Greenes RA, Peterson HE, Protti DJ, eds. *Conference Proceedings of Medinfo 1995*. Amsterdam: North-Holland Publishing Company, 1995;8(Pt 2):1034.
17. Spencer HC. Perspectives of public health and prevention. *Am J Med Sci* 1995;310(1 Suppl):S83-S85.
18. Wald JS, Rind D, Safran C, Kowaloff H, Barker R, Slack WV. Patient entries in the electronic medical record: An interactive interview used in primary care. In: Gardner RM, ed. *Conference Proceedings of the Annual Symposium on Computer Applications in Medical Care 1995*. Philadelphia: Hanley and Belfus, Inc., 1995:147-151.
19. Makoul G, Arntson P, Schofield T. Health promotion in primary care - Physician-patient communication and decision making about prescription medications. *Soc Sci Med* 1995;41(9):1241-1254.
20. Pringle M, Ward P, Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. *Br J Gen Pract* 1995;45(399):537-541.
21. Pories WJ. Is the medical record dangerous to our health? *North Carolina Med J* 1990;51(1):47-55.
22. Gillies J, Barry D, Crane J, Jones D, MacLennan L, Pearce N, Reid J, Toop L. A community trial of a written self management plan for children with asthma. *N Z Med J* 1996;109(1015):30-33.
23. Weed LL, Wakefield JS, Yarnall SR. *Implementing the problem-oriented medical record*. Seattle: Medical Computer Services Association, 1976.
24. Ball MJ, Collen MF. *Aspects of the computer-based patient record*. New York: Springer Verlag, 1992.
25. Duggan AK, Starfield B, DeAngelis C. Structured encounter form: The impact on provider performance and recording of well-child care. *Pediatrics* 1990;85(1):104-113.
26. Tange HJ. *Medical Narratives in the Electronic Medical Record. Towards a searching structure with optimal granularity*. Thesis 1997.
27. Geiger G, Merrilees K, Walo R, Gordon D, Kunov H. An analysis of the paper-based health record: Information content and its implications for electronic patient records. In: Greenes RA, Peterson HE, Protti DJ, eds. *Conference Proceedings of Medinfo 1995*. Amsterdam: North-Holland Publishing Company, 1995;8(Pt 1):295.
28. Hohnloser JH, Purner F, Kadlec P. Coding medical concepts: A controlled experiment with a computerised coding tool. *Int J Clin Monit Comput* 1995;12(3):141-145.
29. Weingarten SR, Stone E, Green A, Pelter M, Nessim S, Huang H, Kristopaitis R. A study of patient satisfaction and adherence to preventive care practice guidelines. *Am J Med* 1995;99(6):590-596.
30. Hammermeister KE, Shroyer AL, Sethi GK, Grover FL. Why it is important to demonstrate linkages between outcomes of care and processes and structures of care. *Med Care* 1995;33(10 Suppl):OS5-OS16.
31. Hobbs FD, Hawker A. Computerised data collection: Practicability and quality in selected general practices. *Fam Pract* 1995;12(2):221-226.
32. IOM committee. *The IOM report and public health*. *J Am Med Ass* 1990;264(4):508-509.
33. Dornan S, Murray FE, White G, McGilchrist MM, Evans JM, McDewitt DG, MacDonald TM. An audit of the accuracy of upper gastrointestinal diagnoses in Scottish Morbidity Record 1 data in

- Tayside. *Health Bull Edinb* 1995;53(5):274-279.
34. Tufo HM, Speidel JJ. Problems with medical records. *Med Care* 1971;9(6):509-517.
35. Galfalvy HC, Reddy SM, Niewiadomska-Bugaj M, Friedman S, Merkin B. Evaluation of Community Care Network (CCN) system in a rural health care setting. In: Gardner RM, ed. *Conference Proceedings of the Annual Symposium on Computer Applications in Medical Care 1995*. Philadelphia: Hanley and Belfus, Inc., 1995:698-702.
36. Burnum JF. The misinformation era: The fall of the medical record. *Ann Intern Med* 1989;110(6):482-484.
37. Moidu K, Falsone JJ, Nair S. Computer-based patient record: The essential data set approach. In: Ozbolt JG, ed. *Transforming Information, Changing Health Care - Conference Proceedings of the Annual Symposium on Computer Applications in Medical Care 1994*. Philadelphia: Hanley and Belfus, Inc., 1994:491-495.
38. Weed LL. Medical records, medical education, and patient care: The problem-oriented record as a basic tool. Chicago: Year Book Medical Publications, 1971.
39. Grogan S, Conner M, Willits D, Norman P. Development of a questionnaire to measure patients' satisfaction with general practitioners' services. *Br J Gen Pract* 1995;45(399):525-529.
40. Lamberts H, Hofmans-Okkes I. Episode of care: A core concept in family practice. *J Fam Pract* 1996;42(2):161-169.
41. Selzer R, Bonomo Y, Patton G. Primary care assessment of a patient with an eating disorder. *Aust Fam Physician* 1995;24(11):2032-2036.
42. Swanson JA, Hoecker JL. Concise review for primary-care physicians. *Mayo Clin Proc* 1996;71(2):179-183.
43. Callahan CM, Tierney WM. Health services use and mortality among older primary care patients with alcoholism. *J Am Geriatr Soc* 1995;43(12):1378-1383.
44. Goldacre MJ, Ferguson JA. In-patient workload in medical specialties: 1. Demographic profiles and time trends from linked statistics. *Quart J Med* 1995;88(9):649-659.
45. Petrie JC, McIntyre N. *The Problem Oriented Medical Record (POMR): Its use in hospitals, general practice and medical education*. Edinburgh: Churchill Livingstone, 1979.
46. Berni R, Readey H. Problem-oriented medical record implementation: Allied health peer review. Saint Louis: Mosby, 1974.
47. Boulton C, Pacala JT, Boulton LB. Targeting elders for geriatric evaluation and management: Reliability, validity, and practicality of a questionnaire. *Aging (Milano)* 1995;7(3):159-164.
48. McDonald CJ, Tierney WM. Computer-stored medical records - Their future role in medical practice. *J Am Med Ass* 1988;259(23):3433-3440.
49. Starfield B. Primary care and health - A cross-national comparison. *J Am Med Ass* 1992;268(15):2032-2033.
50. *Nederlands Huisartsen Genootschap referentiemodel 1985*. Utrecht: WCIA, 1985.
51. *Nederlands Huisartsen Genootschap referentiemodel 1995*. eds. JJ v Overbeeke en HP Westerhof. Utrecht: WCIA, 1996.
52. Safran C, Rind D, Citroen M, Bakker AR, Slack WV, Bleich HL. Protection of confidentiality in the computer-based patient record. *MD Comput* 1995;12(3):187-192.
53. O'Brien JA. *Introduction to information systems in business management*. 6th ed. Homewood: Richard D. Irwin, 1991.
54. Stoner JAF, Freeman RE. *Management*. 5th ed. Englewood Cliffs: Prentice-Hall, Inc., 1992.
55. Davis GB, Olson MH. *Management information systems: Conceptual foundations, structure, and development*. 2nd ed. McGraw-Hill Series in Management Information Systems. New York: McGraw-Hill, Inc., 1985.
56. Sprague Jr. RH, Carlson ED. *Building effective decision support systems*. Englewood Cliffs: Prentice Hall, Inc., 1982.
57. Sprague Jr. RH, Watson HJ. *Decision support systems - Putting theory into practice*. Englewood Cliffs: Prentice-Hall International, Inc., 1986.
58. Turban E. *Decision support and expert systems (Managerial perspectives)*. New York: Macmillan Publishing Company, 1988.

59. Alter SL. Decision support systems: Current practice and continuing challenges. Addison-Wesley Series on Decision Support. Reading: Addison-Wesley Publishing Company, 1980.
60. Bennett JL. Building decision support systems. Addison-Wesley Series on Decision Support. Reading: Addison-Wesley Publishing Company, 1983.
61. Keen PGW, Scott Morton MS. Decision support systems - An organizational perspective. Addison-Wesley Series on Decision Support. Reading: Addison-Wesley Publishing Company, 1978.
62. Mclean E, Sol HG. DSS: A decade in perspective. Amsterdam: North-Holland Publishing Company, 1986.
63. Bodily SE. Modern decision making - A guide to modelling with DSS. New York: McGraw-Hill, Inc., 1985.
64. Farley AM. Diagnostic mechanism modelling. *Int J Patt Recogn AI* 1989;3(2):237-259.
65. Remus W, Kottemann JE. Semi-structured recurring decisions: An experimental study of decision making models and some suggestions for DSS. *MIS Quart* 1987;(June):232-243.
66. Bergamaschi S, Lodi S, Sartori C. Entity-situation: A model for knowledge representation. Bologna: University of Bologna, December 1989; Report No. 66.
67. Methlie LB, Sprague RH. Knowledge representation for DSS. Amsterdam: North-Holland Publishing Company, 1985.
68. Reitsma RF. Inference trees: A method for modelling relational decision rules. Vienna: Conference on Intelligence and Society, 1988.
69. Deutsch T, Carson E, Ludwig E. Dealing with medical knowledge - Computers in clinical decision making. New York: Plenum Press, 1994.
70. American Society for Testing and Materials, E 1460 Standard Specification for Defining And Sharing Modular Health Knowledge Bases (Arden Syntax for Medical Logic Modules), pp. 539-587 (ASTM, Philadelphia, 1992).
71. Hripsak G, Clayton PD, Pryor A, Haug P, Wigertz OB, van der Lei J. The Arden Syntax for Medical Logic Modules. In: Miller RA, ed. SCAMC Proceedings 1990. Amsterdam: Elsevier Publishers, 1990:200-204.
72. Bonnet A. L'Intelligence artificielle - Promesses et Réalités. Paris: InterÉditions, 1984.
73. Goldenthal N. Expert systems and artificial intelligence. Chesterland: Weber Systems, Inc., 1987.
74. Blum BI, Duncan K, eds. A history of medical informatics. ACM Press History Series. Reading: Addison-Wesley Publishing Company, 1990.
75. Zahlmann G, Franczykova M, Henning G, Strube M, Hüttl I, Hummel I, Bruns W. DIABETEX - A decision support system for therapy of type I diabetic patients. *Comp Meth Prog Biomed* 1990;32(3-4):297-301.
76. Middleton B, Shwe MA, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, Cooper GF. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR Knowledge Base - II. Evaluation of diagnostic performance. *Meth Inform Med* 1991;30:256-267.
77. Shwe MA, Middleton B, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, Cooper GF. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR Knowledge Base - I. The probabilistic model and inference algorithms. *Meth Inform Med* 1991;30:241-255.
78. Rutscher A, Salzsieder E, Thierbach U, Fischer U, Albrecht G. KADIS - A Computer-aided decision support system for improving the management of type-I diabetes. *Experim and Clin Endocrin* 1990;95(1):137-147.
79. Buchanan BG, Shortliffe EH, eds. Rule-based expert systems - The MYCIN experiments of the Stanford Heuristic Programming Project. The Addison-Wesley Series in Artificial Intelligence. Reading: Addison-Wesley Publishing Company, 1984.
80. Masarie FE, Miller RA. Quick Medical Reference (QMR): A microcomputer based diagnostic support tool for internal medicine. In: Barber B, Cao D, Qin D, Wagner G, eds. Conference Proceedings of Medinfo 1989. Amsterdam: North-Holland Publishing Company, 1989;6(Pt 2):1209.
81. Vissers M. Diagnostic and Therapeutic Protocols in Medical Practice. Thesis 1996.
82. Cannon SR, Gardner RM. Experience with a computerized interactive protocol system using HELP. *Comp Biomed Res* 1980;13:399-409.

83. McDonald CJ, Hui SL, Smith DM, Tierney WM, Cohen SJ, Weinberger M, McCabe GP. Reminders to physicians from an introspective computer medical record. *Ann Intern Med* 1984;100:130-138.
84. Lei vd J, Does vd E, Man in 't Veld AJ, Musen MA, Bommel v JH. Response of general practitioners to computer-generated critiques of hypertension therapy. *Meth Inform Med* 1993;32(2):146-153.
85. Mackintosh CA, Bradley P. A management information system for general practice. In: Richards B, MacOwan H, eds. *Conference Proceedings HC 93 - Current Perspectives in Healthcare Computing 1993*. Weybridge: BJHC Books, 1993:559-566.
86. Crombie JK, Davies HJO. Computers and audit: servants or sirens? *BMJ* 1991;303:853.
87. Linnarsson R, Nordgren K. A shared computer-based problem-oriented patient record for the primary care team. In: Greenes RA, Peterson HE, Protti DJ, eds. *Conference Proceedings of Medinfo 1995*. Amsterdam: North-Holland Publishing Company, 1995;8(Pt 2):1663.
88. Wyatt J, Spiegelhalter D. Evaluating medical expert systems: What to test, and how? In: Talmon JL, Fox J, eds. *Knowledge based systems in medicine: Methods, applications and evaluation*. Berlin: Springer Verlag, 1989;47:274-289.
89. O'Moore R. The conception of a medical computer system. In: van Gennip EMSJ, Talmon JL, eds. *Assessment and evaluation of information technologies in medicine*. Amsterdam: IOS Press, 1995;17:45-49.
90. O'Moore R., Engelbrecht R. The evaluation of medical decision support and expert systems: Reflections on the literature. In: Talmon JL, Fox J, eds. *Knowledge based systems in medicine: Methods, applications and evaluation*. Berlin: Springer Verlag, 1989;47:263-273.
91. Grémy F, Bonnin M. Evaluation of automatic health information systems - What and how? In: van Gennip EMSJ, Talmon JL, eds. *Assessment and evaluation of information technologies in medicine*. Amsterdam: IOS Press, 1995;17:9-20.
92. Engelbrecht R, Rector A, Moser W. Verification and validation. In: van Gennip EMSJ, Talmon JL, eds. *Assessment and evaluation of information technologies in medicine*. Amsterdam: IOS Press, 1995;17:51-66.
93. Smeets RPAM. *Decisions and Strategies in Drug Treatment of Epilepsy*. Thesis 1996.
94. Bankowitz RA, Lave JR, McNeil MA. A method for assessing the impact of a computer-based decision support system on health care outcomes. *Meth Inform Med* 1992;31(1):3-11.
95. Bjerregaard B, Brynitz S, Holst-Christensen J, Kalaja E, Lund-Christensen J, Hilden J, de Dombal FT, Horrocks JC. Computer-aided diagnosis of the acute abdomen: A system from Leeds used on Copenhagen patients. In: de Dombal FT, Grémy F, eds. *Decision Making and Medical Care*. New York: North-Holland Publishing Company, 1976:165-174.
96. Kolodner JL. Improving human decision making through case-based decision aiding. *AI Mag* 1991;12(Summer):52-68.
97. Andrus CH, Daly JL. Evaluation of surgical services in a large university-affiliated VA hospital: Use of an in-house-generated quality assurance data base. *South Med J* 1991;84(12):1447-1450.
98. Barnett GO, Winickoff RN. Quality assurance and computer-based patient records. *Am J Publ Health* 1990;80(5 May):527-528.
99. Berg M. *Rationalizing medical work: Decision support techniques and medical practices*. Cambridge: MIT Press, 1997.
100. Boran G, O'Moore R, Grimson W, Peters M, Hasman A, Groth T, van Merode F. A new clinical laboratory information system architecture from the Openlabs project offering advanced services for laboratory staff and users. *Clinica Chimica Acta* 1996;248:19-30.
101. Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, et al. Performance of four computer-based diagnostic systems. *New Eng J Med* 1994;330:1792-1796.
102. Eisenberg JM. Physician utilization - The state of research about physicians' practice patterns. *Med Care* 1985;23:461-483.
103. Schectman JM, Elinsky EG, Pawlson LG. Effect of education and feedback on thyroid function testing strategies of primary care clinicians. *Arch Intern Med* 1991;151:2163-2166.
104. Martin AR, Wolf MA, Thibodeau LA, Dzau V, Braunwald E. A trial of two strategies to modify

- the test-ordering behavior of medical residents. *N Engl J Med* 1980;303:1330-1336.
105. Fowkes FGR, Evans KT, Hartley G, et al. Multicentre trial of four strategies to reduce use of a radiological test. *Lancet* 1986;i:367-370.
106. Marton KI, Tul V, Sox HC. Modifying test-ordering behavior in the outpatient medical clinic - A controlled trial of two educational interventions. *Arch Intern Med* 1985;145:816-821.
107. Winkens RAG, Pop P, Bugter-Maessen AMA, Grol RPTM, Kester ADM, Beusmans GHMI, Knottnerus JA. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *The Lancet* 1995;345:498-502.
108. Eisenberg JM. An educational program to modify laboratory use by house staff. *J Med Educ* 1977;52:578-581.
109. Berwick DM, Coltin KL. Feedback reduces test use in a health maintenance organization. *J Am Med Ass* 1986;255:1450-1454.
110. Everett GD, de Bois CS, Chang PF, Holets T. Effect of cost education, cost audits, and faculty chart review on the use of laboratory services. *Arch Intern Med* 1983;143:942-944.
111. Sommers LS, Sholtz R, Shepherd RM, Starkweather DB. Physician involvement in quality assurance. *Med Care* 1984;22:1115-1138.
112. Schroeder SA, Myers LP, McPhee SJ, et al. The failure of physician education as a cost containment strategy. *J Am Med Ass* 1984;252:225-230.
113. Pop P, Winkens RAG. A diagnostic centre for general practitioners: Results of individual feedback on diagnostic actions. *J R Coll Gen Pract* 1989;39:507-508.
114. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992;304:1093-1096.
115. Winkens RAG, Pop P, Bugter-Maessen AMA, Grol RPTM, Kester ADM, Beusmans GHMI, Knottnerus JA. Effect of routine individual feedback over nine years on general practitioners' requests for tests. *Br Med J* 1996;312:490.
116. Dowling PT, Alfonsi G, Brown MI, Culpepper L. An education program to reduce unnecessary laboratory tests by residents. *Academic Med* 1989;64(2):410-412.
117. Drinka PJ, Amberson J, Voeks SK, Schomisch J, Schirz P, Christensen D. Low TSH levels in nursing home residents not taking thyroid hormone. *J Am Geriatr Soc* 1996;44(5):573-577.
118. Moore DC. Natural course of 'subclinical' hypothyroidism in childhood and adolescence. *Arch Pediatr Adolesc Med* March 1996;150:293-297.
119. van der Velden J, de Bakker DH, Claessens AAMC, Schellevis FG. Een nationale studie naar ziekten en verrichtingen in de huisartspraktijk - Basisrapport: Morbiditeit in de huisartspraktijk. Utrecht: Nederlands Instituut Voor onderzoek van de EersteLijnsgezondheidszorg (NIVEL), August 1991.
120. Hart IR, Newton RW, eds. The new medicine - An integrated system of study - Volume 2: Endocrinology. Lancaster: MTP Press Limited, 1983:2-3.
121. Brombacher PJ, Marell GJ, Westerhuis LW. Laboratory work flow analysis and introduction of a multi-functional analyser. *Eur J Clin Chem Clin Biochem* 1996;34(3):287-292.
122. Gend van JM, Pelt van J, Cleef TH, Mangnus TM, Muris JW. Kwaliteitsverbeteringsproject 'laboratoriumdiagnostiek door huisartsen' leidt tot aanzienlijke reductie van het aantal laboratoriumanalyses. *Ned Tijdschr Geneesk* 1996;140(9):495-500.
123. Rubenstein LV, McCoy JM, Cope DW, Barrett PA, Hirsch SH, Messer KS, Young RT. Improving patient quality of life with feedback to physicians about functional status. *J Gen Intern Med* 1995;10(11):607-614.
124. Elson RB, Connelly DP. Computerized patient records in primary care. Their role in mediating guideline-driven physician behaviour change. *Arch Fam Med* 1995;4(8):698-705.
125. Almeida CA, Diez FA, Doctorovich S, Iusoff O, Mollerach FE, Yabo R. Indice diagnostico de hipertiroidismo. *La Prensa Medica Argentina* 15 junio 1969;56(15):659-662.
126. Murray IPC. The clinical diagnosis of thyroid disease. *Med J Austral* 30 May 1964;51(22):827-831.
127. Viherkoski M. The clinical index in hyperthyroidism. *Duodecim* 1971;87(11):890-895.
128. Kálmán K, Leövey A, Szabó T. Clinical and immunological follow-up of pregnant women with

- Graves' Disease. *Act Med Hung* 1986;43(1):3-11.
129. Edling C. Kliniskt index och laboratoriedata vid diagnos. *Lakartidningen* 1970;67(4):370-374.
  130. Crooks J, Murray IPC, Wayne EJ. Statistical methods applied to the clinical diagnosis of thyrotoxicosis. *Quart J Med* April 1959;XXVIII(110):211-234.
  131. Crooks J, Murray IPC, Wayne EJ. Basal metabolic rate in thyrotoxicosis. *Lancet* 1958;22 March:604-607.
  132. Wayne EJ. Clinical and metabolic studies in thyroid disease. *Br Med J* 1960;2 January(5165):1-11.
  133. Wayne EJ. The diagnosis of thyrotoxicosis. *Br Med J* 1954;20 February(4859):411-419.
  134. Börner W, Lautsch M, Moll E, Rauh E. Verbesserung der klinischen Hyperthyreosedagnostik mit Hilfe eines einfachen Punktesystems. *Deut Med Woch* 1968;8 November(45):2173-2174.
  135. Gurney C, Griffith O, Hall R, Roth M, Harper M, Smart GA. Newcastle Thyrotoxicosis Index. *Lancet* 1970;19 December:1275-1278.
  136. von Laubinger G. Hyperthyreosedagnostik mit Hilfe eines Punktesystems. *Med Klinik* 1968;63:1708-1709.
  137. Fragu P, Alperovitch A, Patois E. Screening and diagnosis of hyperthyroidism: An attempt at test reduction. *J Epid Com Health* 1979;33:215-218.
  138. Simonin R, Sambuc R, San Marco JL, Gavelle-Desbrosses C. Un nouvel index de thyrotoxicose (Base de calcul et applications). *Pathologie Biologie, Société Française de Biologie Médicale* 1982;30(8):694-702.
  139. Simonin R, Miller G, Leandri M, Simeoni P. Le diagnostic clinique des hyperthyreoses (La notion d'index thyroïdien). *Archives de Médecine Générale et Tropicale* 1960;37:75-85.
  140. Möpert S, Vollmar R. Rechtfertigt ein Hyperthyreose-Symptomkomplex-Punktesystem eine Ablehnung weiterer Schilddrüsendiagnostik? *Radiobiologia, Radiotherapia: Int Zeitschr für Strahlenther, Strahlenbiol, Strahlenphys, Nuklearned* 1970;11(2):179-187.
  141. Wayne EJ. Clinical and metabolic studies in thyroid disease. *Br Med J* 1960;9 January:78-90.
  142. Billewicz WZ, Chapman RS, Crooks J, Day ME, Gossage J, Wayne E, Young JA. Statistical methods applied to the diagnosis of hypothyroidism. *Quart J Med* 1969;April:255-266.
  143. Seshadi MS, Samuel BU, Kanagasabapathy AS, Cherian AM. Clinical scoring system for hypothyroidism: Is it useful? *J Gen Intern Med* 1989;4(6):490-492.
  144. Wiersinga WM. Over het ontstaan van schildklierziekten. Amsterdam: Amsterdam University Press, 1995.
  145. Caldwell G, Gow SM, Sweeting VM, Kellett HA, Beckett GJ, Seth J, Toft AD. A new strategy for thyroid function testing. *Lancet* 1985;1:1117-1119.
  146. Klee GG, Hay ID. Assessment of sensitive thyrotropin assays for an expanded role in thyroid function testing: Proposed criteria for analytic performance and clinical utility. *J Clin Endocrin Metab* 1987;64(3):461-471.
  147. Hall R, Besser M. Fundamentals of clinical endocrinology. New York: Longman Publishers, 1989.
  148. Hennemann G, Docter R, Vos RA, van Toor H, Krenning EP. Vergelijking van gebruikelijke schildklierfunctietests met nieuwere TSH-tests bij patiënten met ziekten die niet de schildklier betreffen. *Ned Tijdschr Geneesk* 1987;131(51):2355-2359.
  149. Mudde AH, Bastiaanse AJ, Jonkers H. TSH-IRMA als eerste diagnosticum bij vermoeden van hyperthyreoidie. *Ned Tijdschr Geneesk* 1987;131(51):2364-2368.
  150. Nuutila P, Irjala K, Viikari J, Prinssi V-P, Kaihola H-L. Comparative evaluation of serum thyroxine, free thyroxine and thyrotropin determinations in screening of thyroid function. *Ann Clin Res* 1988;20:158-163.
  151. Hay ID, Klee GG. Thyroid dysfunction. *Endocrin Metab Clinics North Am* 1988;17(3):473-509.
  152. Ericsson U-B, Fernlund P, Thorell JI. Evaluation of the usefulness of a sensitive immunoradiometric assay for thyroid stimulating hormone as a first-line thyroid function test in an unselected patient population. *Scand J Clin Lab Invest* 1987;47:215-221.
  153. Coli A, Ferdeghini M, Malvano R, Madeddu G, Prontera C, Scarpato F. Analytical and clinical assessment of a commercial kit for the simultaneous determination of free thyroxine and thyrotropin. *J Nucl Biol Med* 1991;35(1):24-32.



## **Chapter 3**

### **Computer-based assistance in family medicine**

The article presented in this chapter has been accepted for publication in

Dupuits FMHM, Hasman A, Pop P. Computer-based assistance in primary care.  
Computer Methods and Programs in Biomedicine 1997; in press.

The references <sup>1,2,9-22,24,26,33,34</sup> included in the List of Publications also concern the research described here.



## ABSTRACT

In this paper, the fully integrated General Practitioner (GP) Information and Research System HIOS and Decision Support System HIOS+ are described. Both systems were developed with the Phased Development Methodology (PDM), which is also presented. HIOS consists of an administrative module, medical modules, and a research module to support GPs in daily practice activities and research. It was implemented at a general practice and a health centre, and was evaluated positively. HIOS+ was developed by extending HIOS with modules which offer several types of decision support to assist GPs in (diagnostic) decision making. The modules offer passive and active decision support. HIOS+ can present textbook information, generate advice and warnings, support the creation of models, generate reports, and can be used in education. Some HIOS+ modules were evaluated by GPs in evaluation studies, all with positive results.

## KEYWORDS

Decision Support System, Data Base, Model Base, Design and Development Technique, Medical Diagnosis, Family Medicine

## Introduction

General Practitioners (GPs) want of their own accord or are stimulated by others to provide quality of care. To accomplish this, a high data quality is needed. Many data are stored in patient records at general practices. Since these data usually are collected and stored in an unstructured way, they can not easily be utilized in research or decision support. To perform epidemiological studies and to improve general practice management and the quality of care, GPs need tools which support structured data capture, data storage, data interpretation, and data retrieval. GP information systems are such tools.

General practices are well suited for computerization<sup>1,2</sup>. Much time and effort has been put into the development of GP information systems, whose main function is to support GPs in their daily activities. Several computer applications and their impact on primary care are described in literature<sup>3-11</sup>. Already in 1985, the Department of Medical Informatics of the Maastricht University (formerly called the University of Limburg) started to investigate GPs' requirements concerning information systems. The major aim of the study was to develop a GP Information and Research System, called HIOS, which would meet the system requirements indicated by GPs. The GPs organized in the 'Registration Network General Practices' of the university's Department of Family Medicine participated in this study. Objectives were to provide computer-based facilities

- for:
- the registration of administrative data;
  - the registration of medical data; and
  - data analysis for medical and research purposes.

Software development was based partly on own research results and partly on the 1985 reference guide of the Dutch College of General Practitioners<sup>12</sup>, which at that time stipulated guidelines for administrative patient data modules but discussed medical and other modules only briefly.

In 1989, the focus of the research was changed to computer-based decision support for GPs. In a survey, namely, GPs had indicated the need for active and passive computer-based support concerning especially diagnostic decision making in daily practice. The major aim of the research was to develop and evaluate computer-based decision support for GPs. Objectives were to provide computer-based:

- passive support (to present textbook information);
- active data-oriented support (to establish a diagnosis on the basis of patient data);
- active goal-oriented support (to check whether an entered diagnosis fulfils the corresponding ICHPPC-II-Defined criteria);
- tools for generating and communicating reports; and
- decision support in the training of medical professionals and students.

From 1989 onwards, several decision support modules were developed and integrated with HIOS into the Decision Support System (DSS) HIOS+. Because of this integration, the decision support modules would be able to immediately access all stored data.

In this paper, methods and techniques used in as well as results of the development and the evaluation of HIOS and HIOS+ are described and discussed.

## **Methods and techniques**

### *The development methodology*

The in-house established Phased Development Methodology (PDM)<sup>13</sup> was used in software development and provided software engineers with methods, techniques, and tools. The philosophy behind PDM is that a system can only be successfully designed and developed when users are involved during its life cycle. Moreover, it is a structured approach; both the analysis and the design phase focus on the decomposition of various tasks and corresponding software modules. It is also an evolutionary approach in that it adopts a prototyping technique in which each prototype is gradually updated. Four development phases are distinguished.

### User requirements

The first phase involves an inventory and analysis of user requirements. Participating GPs are interviewed individually. In a plenary meeting, the interview results

are discussed and priorities are determined for each need. These needs and priorities are stored on a checklist.

#### Functional and technical specifications

The second phase involves the application of a prototyping technique to obtain the system specifications needed to fulfil the user requirements. The starting point in the development process is the checklist. During the development, a new system version is obtained by extending the current one on the basis of information on the checklist.

Design problems are analysed using a top-down/bottom-up analysis technique. The developer's and the GPs' scopes are changed several times from bottom up to top down or vice versa during the analysis of a problem. This ensures an iterative process. Simon's decision-making model<sup>14</sup>, furthermore, forms the basis for modelling decisions with respect to the development process. The functional and technical specifications are obtained and implemented using an evolutionary prototyping approach.

#### The final system

The third phase involves the implementation of the final system. This is done after all entries on the checklist are realized or the remaining functionalities are considered not feasible in an economical, practical, or technological sense. The routine monitoring concerning entries on the checklist, system deficiencies, technological progress, and changed or additional user demands continues in order to be able to adapt the checklist.

#### The software evaluation

In the fourth phase, the systems are demonstrated to GPs. The systems are tested both in laboratory test situations and in daily practice (field tests). The GPs are interviewed or asked to complete questionnaires.

### *Techniques*

#### Passive support

In passive support, use is made of a hypertext-like approach. Topics on which textbook information is available are highlighted either by colour or by underlining. All the topics together with the corresponding textbook information are stored in a data base. The topics are linked, where appropriate.

#### Active support

Several techniques are applied to implement protocols, rule-based systems, scoring systems, and decision trees. In addition, prespecified queries are used to check entered diagnostic codes against data in the data base.

The architecture for data-oriented support is a kind of blackboard system. In such

a system, several programs, called Knowledge Sources (KSs), are available, each of which supports the diagnostic process with a model. These KSs determine whether they can be activated by triggers. The triggers obtain their information from the blackboard on which each of the KSs writes its output. Triggers are represented as chains of logical expressions consisting of logical operators and items like signs/symptoms, physical examination data, drugs, results of laboratory tests, and diagnoses. Each trigger activates a KS. All necessary patient data are also written onto the blackboard.

The goal-oriented support is based on Medical Logic Modules (MLMs)<sup>15-18</sup>, which incorporate created queries. The Arden Syntax<sup>15,16</sup> is used in combination with a pseudocompiler<sup>19</sup>. A MLM contains a Maintenance, a Library, and a Knowledge part. The Maintenance and Library parts provide administrative data on the MLM's creator and literature sources. The Knowledge part contains the logic (rules) used in decision support.

For evaluation purposes, a shortened version of the Bailey questionnaire<sup>20,21</sup> extended with open questions<sup>22</sup> is used to assess user satisfaction. Bailey identified factors indicative of a system's usefulness. Users express their satisfaction with each factor (presented as a question) by means of a positive or a negative weight. The importance of the factor is also graded.

### *The software environment and architecture*

The language used in all software developments is Turbo Pascal. The systems are implemented in a Personal Computer (PC) environment, either as stand-alone applications or multi-user applications. The PCs are connected in local area networks.

HIOS and HIOS+ contain a data base component, a user interface, and system management software. Additionally, HIOS+ is equipped with a model base component. The HIOS+ components are typical of all DSSs.

## **Results**

### *The development of HIOS and HIOS+*

PDM could successfully be used in the development of HIOS and HIOS+ in that it was fast, efficient, and effective. Much time and effort was spent on gathering the GPs' ideas and wishes and putting these into practice. Based on users' remarks, each system version was updated and extended. When HIOS was developed in 1986, voice and mouse input proved not to be feasible for practical reasons. Due to the diversity of voices, only the voices of a limited number of persons were recognized, which was an unworkable situation in a health centre. Disturbing sounds from the immediate surroundings,

furthermore, occasionally caused the voice controller to randomly generate words. According to the GPs, the act of moving a mouse for data entry was too disturbing during a GP/patient encounter and the mouse with the underlay took too much space physically. The HIOS and HIOS+ interface, which is based on the use of windows, in combination with function keys, however, were well appreciated.

Eventually, a final HIOS version and final versions of all HIOS+ modules were built and installed. A GP information system with incorporated decision support modules was obtained, which is suitable for use in primary care.

### *Classification and coding*

In order to be able to support GPs effectively, a GP information system has to offer a structured way of data registration. HIOS supports GPs during medical data registration and, thus, enables research. In the registration of medical data, the SOAP approach<sup>23</sup> is used which is a problem-oriented way of recording. The acronym stands for: Subjective findings (symptoms stated by a patient), Objective findings (signs found in medical examination), Assessment (diagnosis), and Plan (therapy). In daily work and research, GPs do not experience major problems in the application of this approach. This may be the result of practical training sessions at the university in which GPs are taught how to use GP information systems, problem-oriented registration, and classifications. In HIOS, additionally, several coding instruments are offered. One of these is the International Classification of Primary Care (ICPC)<sup>24-26</sup>, which specifies codes for diagnoses and therapeutic actions and is accepted by the Dutch College of General Practitioners and the World Health Organization. The International Classification of Health Problems in Primary Care (ICHPPC)-II-Defined<sup>27</sup> can be consulted for criteria belonging to the corresponding diagnoses. OCO<sup>28</sup>, a physical examination code with an ICPC-like coding structure, additionally enables the coding of physical examination results and test outcomes. Finally, a drug classification list is available which also has an ICPC-like coding structure.

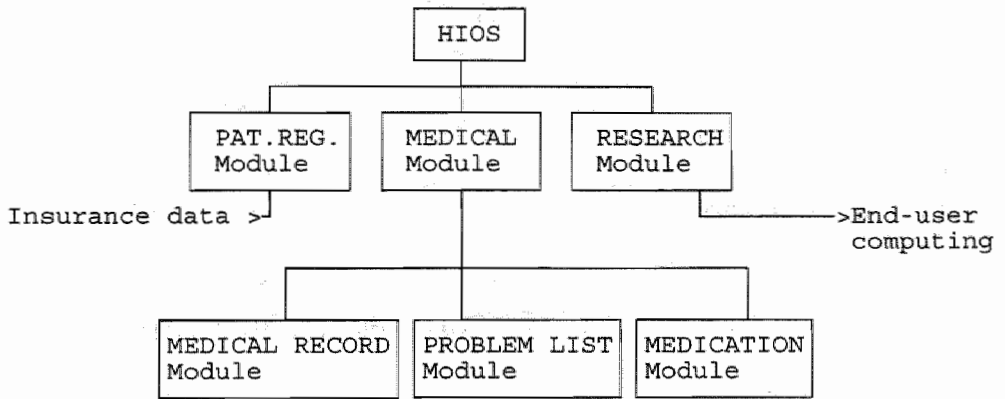
### *The functionality and the structure of HIOS*

HIOS meets the requirements indicated by the interviewed GPs and the GP information system requirements defined by the Dutch College of General Practitioners. In HIOS, not only administrative data (general patient data, billing data, etc.) but also medical data (complaints, problems, diagnoses, etc.) can be stored. It, furthermore, can be used to analyse stored data. HIOS consists of several modules, as is shown in figure 1.

### The Patient Registration Module

This module enables the registration of patient identification data, demographic data, other basic patient facts, and patient background variables like a patient's social status. It can read files containing patient demographic data obtained from sickfund insurance companies. This saves time (especially in case paper-based medical records are converted into computer-based ones) because only data of privately insured patients need to be entered manually into HIOS. To find a patient record, a key is used, containing the first four characters of a patient's surname and six characters of the date of birth.

**Figure 1. The modules of the GP Information and Research System HIOS and the HIOS environment**



The module also supports the processing and the management of data regarding research purposes. Reminders can be generated, and address labels printed. Facilities are offered with which GPs can bill patients periodically or immediately.

### The Medical Module

This module enables the structured registration of medical data by offering coding facilities and the SOAP approach. It contains a Medical Record Module, a Problem List Module, and a Medication Module.

### The Medical Record Module

This module enables the capture of medical data in a userfriendly manner by offering function keys, windows, extensive help functions and search routines, and guarantees data quality by performing numerous data entry checks.

To find a specific ICPC code during registration, a GP is offered several search

options. Firstly, a GP can search for a code by browsing through the alphabetically ordered ICPC codes presented in a window. Secondly, a GP can enter a valid ICPC chapter and/or sequence number. Then, only the ICPC descriptions of that particular chapter and/or sequence number are presented. Thirdly, an entire ICPC code can be entered, resulting in the presentation of the code's corresponding description. Finally, a GP can enter (parts of) words which are then checked against ICPC code descriptions. Matches are shown in a window.

GPs can consult the ICHPPC-II-Defined criteria belonging to an ICPC code when the corresponding diagnosis is considered. The criteria define that particular diagnosis.

The status of a complaint or diagnosis can be registered as old or new. Status data give a GP insight into the evolution (continuation or (renewed) manifestation) of clinical signs/symptoms. Furthermore, a GP's certainty with respect to a registered diagnostic hypothesis can be stored.

By specifying links between entries of different SOAP categories stored during the same encounter, a GP later on can re-establish which symptoms led to which actions and to which diagnoses. A plan or therapy can be traced from a diagnosis. The progression of a patient's disease as a function of time can be monitored by linking current SOAP entries to SOAP entries of a previous encounter or to problem list entries.

Graphical overviews (histograms, pie charts, etc.) can be generated. Developments in e.g. systolic and diastolic blood pressures or growth can be presented graphically.

### The Problem List Module

Chronic and/or urgent problems of a patient can be registered. The state of a problem (active or inactive) and the date of a problem's activation or inactivation can be stored.

It is possible to specify links between problems, and between a problem and SOAP entries in a medical record. Moreover, the state of a problem can be changed from active into inactive and vice versa.

Historical overviews of problems of a selected patient can be made, since the evolution of a problem in time is registered. Using these overviews and specified links between problems and SOAP entries in a medical record, a GP can perform episodic research, i.e. a patient's problems or diseases can be monitored over periods of time. Similar computer applications were described in literature<sup>29-31</sup>.

### The Medication Module

A GP can register data about medications prescribed to a patient. Drug quantities, drug forms, drug dosages, and the number of repeat prescriptions can be stored.

Modifications can be made at all times. Prescription renewal dates are determined automatically.

Links can be specified between prescribed drugs and SOAP entries in a medical record or problems in a problem list. Historical overviews of drug prescriptions can be made, which enable a GP to evaluate and plan medication therapy. The module automatically prints all necessary data.

### The Research Module

Data in the HIOS data base can be used for research by GPs and can be made available to researchers from an affiliated university. An individual GP can make selections in all administrative and medical data stored in the HIOS data base, and can make an ASCII file of anonymized research data (patient privacy) for use by researchers.

End-user computing is supported, i.e. data from the HIOS data base can be exported into an ASCII file and then be further analysed with commercially available software.

### *The functionality and the structure of HIOS+*

As requested by GPs, HIOS+ offers:

- passive support through
  - \* Information Facility (IF);
- active data-oriented support through
  - \* Model Maker (MM) for model creation and
  - \* Advice and Warning (AW) for interactive user support;
- active goal-oriented support through
  - \* Free-text Formalizer (FF) for model (query) creation and
  - \* Formal Interpreter (FI) for interactive user support; and
- additional tools, as there are
  - \* Report Generator (RG) for generating and communicating reports and
  - \* EVALuator (EV) for educative support to medical professionals and students.

### Information Facility (IF)

IF passively supports GPs by presenting requested textbook information on medical subjects using a hypertext-like approach. Upon a GP's request for information about a medical topic, like thyroid disorders, the module presents general information from which again highlighted topics can be selected.

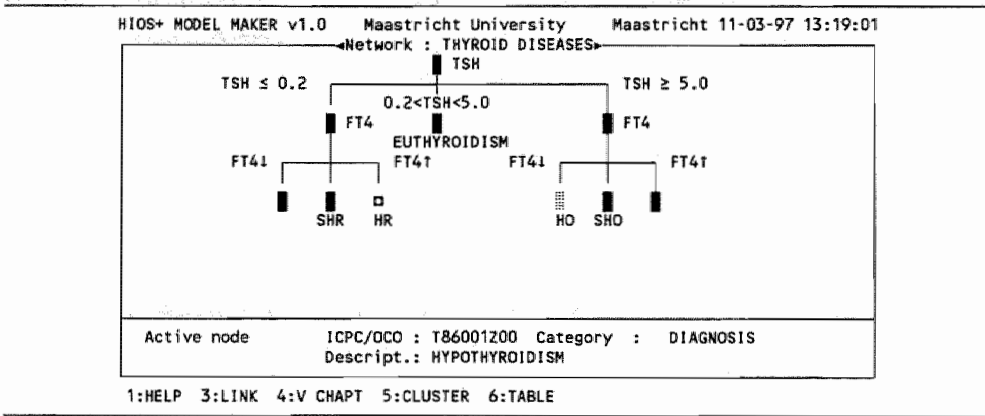


Model Maker (MM)

MM's major function is to enable experts (GPs/consultants) to easily create models and to store these in the model base. The models help to achieve standardization of diagnoses and disease management<sup>32,33</sup>.

Models can be protocols, rules, scoring instruments, and decision trees. For efficiency and consistency reasons, a network structure is used for knowledge representation<sup>34</sup>.

**Figure 2.** Presentation of a model for thyroid diseases (HR = clinical hyperthyroidism, HO = clinical hypothyroidism, SHR = subclinical hyperthyroidism, SHO = subclinical hypothyroidism). The nodes represent items, and the links decision criteria. Laboratory test ranges identify which path is to be followed. The expert last defined the node HO which is a diagnosis with the ICPC/OCO code T86001Z00.



With Model Maker, knowledge is specified which later is used by Advice and Warning (AW). Triggers can be defined which, when satisfied, indicate to AW that the corresponding KS can be executed. With Model Maker, models, like decision trees, protocols, and standards, can be constructed. Figure 2 shows the easy model for diagnosing thyroid diseases based on laboratory tests. Of course, more complex models can be built as well. The nodes in the figure 2 network represent items, and the links decision criteria. The nodes of the network are implemented as records. A record contains various fields, e.g. the name of the node, the decision criteria, and a pointer to the next node which will be advised when a criterion is met. Records presenting end nodes do not contain decision criteria. The sign □ in figure 2 indicates that the information (hyperthyroidism) contained in this end node will trigger another KS leading to a more detailed diagnosis (e.g. Graves' disease, De Quervain's disease). The dotted node is the

last node (in this case clinical hypothyroidism) defined by the expert.

As stated earlier, rules are also represented in the network formalism. If a rule only contains premises connected by 'ANDs', the network reduces to a linear sequence of nodes. If both 'AND' and 'OR' operators are used, a network is obtained. The end nodes are now the possible conclusions of the rules. These conclusions can trigger other KSs.

Model Maker stores the content of each model in the model base.

### Advice and Warning (AW)

Advice and Warning, which is implemented as a kind of blackboard system, generates advice and warnings based on models defined in MM and medical patient data. Advice is defined as any constructive support given by the system to a GP. A warning is defined as any message given by the system to a GP in order to: (i) indicate a violation of an expected condition in a model, or (ii) remind the GP of a previously taken action of which a result is not yet available.

**Figure 3.** Screen illustrating the integration of AW with the GP Information and Research System HIOS, in particular the Medical Record Module. The basic window of the Medical Record Module presents the patient's administrative data along with his medical data of the present encounter. In the overlapping AW window, items of the thyroid illness model are presented about which the GP still has to provide information.

VERSION 210297		HIOS+ ADVICE AND WARNING		11-03-97 08:06:04																			
MEDID10885 Mr. DEMO Medical Informatics		P.O. Box 616 6200 MD Maastricht The Netherlands		Male Single 11 years																			
S   ICPC NR		DESCRIPTION		S P L P.E DATE P.E																			
S1 F01 001 DOUBLE VISION				2																			
S1 Since 2 weeks				2																			
S2 A04 002 PREFERENCE FOR COLD				2																			
O1 T01 001 THYROID MURMUR				2																			
<p>.....</p> <p>→ THYROID ILLNESS FILTER ITEMS →</p> <table border="1"> <tbody> <tr> <td>PULSE RATE &lt; 80/MIN</td> <td>RETRACTION EYELID</td> </tr> <tr> <td>EXOPHTHALMOS</td> <td>EXCESSIVE PERSPIRATION</td> </tr> <tr> <td>FINE FINGER TREMOR</td> <td>DYSPNOEA ON EXERTION</td> </tr> <tr> <td>INCREASED APPETITE</td> <td>BURNING SENSATION IN EYES</td> </tr> <tr> <td>DECREASED APPETITE</td> <td>EYES RED</td> </tr> <tr> <td>HYPERKINESIS</td> <td>PAIN WHILE SWALLOWING</td> </tr> <tr> <td>WEIGHT INCREASED</td> <td>PREFERENCE FOR HEAT</td> </tr> <tr> <td>WEIGHT DECREASED</td> <td>WARM HANDS</td> </tr> <tr> <td>NERVOUSNESS</td> <td>MOIST HANDS</td> </tr> </tbody> </table>						PULSE RATE < 80/MIN	RETRACTION EYELID	EXOPHTHALMOS	EXCESSIVE PERSPIRATION	FINE FINGER TREMOR	DYSPNOEA ON EXERTION	INCREASED APPETITE	BURNING SENSATION IN EYES	DECREASED APPETITE	EYES RED	HYPERKINESIS	PAIN WHILE SWALLOWING	WEIGHT INCREASED	PREFERENCE FOR HEAT	WEIGHT DECREASED	WARM HANDS	NERVOUSNESS	MOIST HANDS
PULSE RATE < 80/MIN	RETRACTION EYELID																						
EXOPHTHALMOS	EXCESSIVE PERSPIRATION																						
FINE FINGER TREMOR	DYSPNOEA ON EXERTION																						
INCREASED APPETITE	BURNING SENSATION IN EYES																						
DECREASED APPETITE	EYES RED																						
HYPERKINESIS	PAIN WHILE SWALLOWING																						
WEIGHT INCREASED	PREFERENCE FOR HEAT																						
WEIGHT DECREASED	WARM HANDS																						
NERVOUSNESS	MOIST HANDS																						
1:HELP 2:SEARCH 3:IF 4:PRLST 6:MEDIC 7:FREE TXT 8:ADVICE 10:OTHER																							

Advice and Warning can generate advice and warnings automatically (the watchdog function) or on request. Firstly, when the trigger of a specific model has fired,

the system automatically activates the corresponding KS. Some KSs generate advice or warnings. Other KSs put intermediate results on the blackboard. Advice or warnings are presented as a blinking signal on the screen. A GP reads the advice or warning (while working in the Medical Record Module) and can act accordingly or ignore the advice. The user can obtain an explanation concerning the advice or warning. Secondly, during data entry, a GP can request diagnostic advice by activating a KS. In figure 3, a GP has activated a scoring system and AW presents the items about which the GP still has to provide information.

When the actual reasoning process is initiated, data about the patient in question are downloaded from the computer-based patient record and stored onto the blackboard. If, finally, a diagnosis is established, all inferred patient data are written to the patient record. If the KS can not be executed further, e.g. if it concerns a protocol and a test is advised which still has to be performed, the data on the blackboard together with the status of the KS are saved to be reactivated when the test outcome becomes available.

#### Free-text Formalizer (FF)

With FF, free text of ICHPPC-II-Defined criteria belonging to diagnostic ICPC codes can be formalized into queries (rules incorporated in MLMs) suited for evaluation by a DSS. The rules are called queries because they are used to query the patient data base. The result of a query is true, false, or unknown, based on the values of available data.

An expert can identify items (single or compound words) in the free-text criteria, link each item to ICPC codes, create a query using the selected items and logical/functional operators (a subset of the Arden Syntax<sup>35</sup>), and check the correctness of the query using the module's pseudocompiler<sup>19</sup>. The validated query is stored in the model base.

#### Formal Interpreter (FI)

FI is integrated in the HIOS Medical Record Module. Its purpose is to query the patient data base, to check whether criteria are met, and to generate advice and/or warnings.

When a GP enters an ICPC code, Formal Interpreter checks whether the criteria for this ICPC code are satisfied. If the patient record does not contain all necessary data, the GP is asked to provide the missing information. The GP can enter the requested information or overrule the system. A warning is generated if the GP enters codes of which the criteria are not satisfied.

Since a query can contain several diagnoses, it is possible that more queries are

executed additionally when validating an entered diagnosis. Acute upper respiratory tract infection (R74000), for instance, is diagnosed if cold (R74009), nasopharyngitis (R74017), pharyngitis (R74019), and rhinitis (R74024) are present (inclusions) and influenza (R80000) is absent (exclusion). Inclusions must be true and exclusions false in order to corroborate the initial query. Whenever the query belonging to R74000 is processed, the queries belonging to the other codes are also evaluated.

#### Report Generator (RG)

RG enables a user to design referral letters. With RG, a GP can draft a letter using data stored in the HIOS data base, and exchange this electronically with other users by means of commercially available electronic data interchange software. All or part of a patient medical history, laboratory test requests, and reasons for test requests or referrals, e.g., can be sent to a consultant. In turn, a consultant can return discharge messages or laboratory test results to a GP, who can then store the necessary information into the GP information system.

#### Evaluator (EV)

EV enables (future) medical professionals to evaluate their knowledge about certain diseases. Evaluator presents a diagnosis of a patient and then asks a user to indicate which information is needed to corroborate the disease. Possibly relevant items are presented in a list. Those items have to be selected which, according to the user, have to be present or absent to establish the diagnosis. If a laboratory test is needed to determine the disorder, the user is asked to indicate in which range of values the outcome should lie, given the disease. The range is presented graphically. Feedback is given to a user about his/her degree of knowledge concerning presented diseases.

#### *The software evaluation*

HIOS and HIOS+ were evaluated during development to assess several aspects of these systems. The factors defined by Bailey and the open questions were evaluated. An important aspect is userfriendliness, since this factor plays an important role in system acceptance. According to the users, HIOS and HIOS+ are userfriendly because they offer help windows, are easy to use, have a high degree of consistency and transparency in the use of function keys, and offer graphic displays. HIOS and HIOS+ can carry out functions properly in small as well as large configurations. Response times are acceptable in both configurations. Feedback to users is presented in an eye-catching way. HIOS+ presents messages (advice and warnings) as blinking signals. HIOS and HIOS+ are highly flexible information systems because of the many ways in which stored data can be

analysed. Finally, an important aspect is the integration of a DSS with a registration system. In HIOS+, decision support modules are integrated with HIOS. This prevents duplication of data and ensures the practical use of the systems by GPs.

HIOS was implemented in a health centre of 6 GPs and about 12,000 patients and in a rural practice of 2 GPs and about 3,200 patients. It was used in daily practice and additionally supported the conversion of paper-based medical records into computer-based ones by its ability to automatically incorporate sickfund data provided by two Dutch sickfund insurance companies. Only data of privately insured patients had to be entered manually from the paper medical record. The conversion process and data validation took about 5 months at either implementation site. HIOS was successfully used by the GPs. In an evaluation of the system by its users, the GPs indicated that HIOS was reliable and fast, and that it offered more functionality than was required in the reference guide of the Dutch College of General Practitioners.

Several decision support modules of HIOS+ were implemented in a general practice of 3 GPs and about 3,200 patients. All of these modules were positively evaluated by their users. The GPs indicated that the modules were fast and that system performance was high. They especially appreciated the active and passive decision support functions. The evaluation results of Free-text Formalizer<sup>36</sup>, Formal Interpreter<sup>22,36</sup>, and EVALuator<sup>37</sup> were published in literature.

## Discussion

### *Advantages of the HIOS+ structure*

The use of a model base and a data base separately from the software makes HIOS+ very flexible. This structure enables experts to individually create diagnostic and therapeutic models without any assistance from the developer.

A recent study in which two browsers were evaluated<sup>38</sup> showed that users are not able to always specify a query correctly. Therefore, experts should preassemble e.g. queries, and users should select desired queries from a list. HIOS+ also makes use of precompiled queries.

### *Problems in modelling*

Free-text criteria often were not structured enough for model creation, or medical knowledge was not available at all.

Texts for instance appeared not to be detailed enough due to terms like "characteristic picture/appearance", "acute onset of", or "atypical lymphocytes". This

made them difficult to formalize unless the terms were used themselves. In that case, however, the possibility existed that GPs used other terms during patient data entry. Such texts, therefore, needed to be described in terms which were stored in the medical record. In that case, all possible synonyms were taken care of, although this was problematic.

No or only little information was found in literature or textbooks about e.g. pretest probabilities of signs, symptoms, and physical examination results. This makes the use of Bayes' theorem for inferring diagnoses problematic<sup>39</sup>.

### *Privacy*

Because of the sensitive nature of the data, arrangements were made to secure patient privacy. Administrative and medical staff had their own access rights to read, modify, or store information concerning their daily activities in a health centre or a GP practice. Use was made of usernames and passwords. Printing facilities were centralized in secured environments. Data for research purposes were anonymized. Only GPs providing the anonymized patient data could be asked to supply additional information. A request was always evaluated by a GP keeping in mind a patient's privacy.

### *A DSS comparison*

A comparison of HIOS+ with four other DSSs which are successfully applied in primary care indicates commonalities.

DERMIS<sup>40</sup> is a computerized skin disease diagnostic system which produces a list of diagnoses based on probabilities calculated using Bayes' theorem. DERMIS placed the correct diagnosis first in 76% and within the first three suggestions in 95% of the occasions. Like HIOS+, it runs on a PC, allows data entry during an encounter, and offers arrangements for performing clinical trials in general practice.

The Physician's Desk Reference and Database (PDRD)<sup>41</sup> is a rule-based diagnostic DSS for primary care concerning a broad spectrum of general diseases. It is currently being assessed by its users. Like HIOS+, it offers facilities for the registration of both administrative data and medical data in daily practice, it is a rule-based diagnostic system, and it runs on a PC.

HIOS+ was compared with systems described by Elson et al.<sup>42</sup>. These systems provide both passive and active decision support. Diagnostic decision support is given on a broad spectrum of diseases. Elson et al. use the same definitions as are described in this article. The rule-based systems improved GP guideline compliance, generally by reducing oversight or error. Their applicability in primary care, however, was limited by the

incompleteness and inaccuracies of the knowledge bases. The use of appropriate standards and classifications, as is the case in HIOS+, therefore, is considered essential by Elson et al.

SWDESTAR<sup>43</sup> is a diagnostic DSS consisting of a vocabulary (for the identification of terms which are used for registration purposes), a patient data base, and a knowledge base concerning a broad spectrum of diseases. According to the SWDESTAR developers, integration with a computer-based patient record system is the key issue for a successful introduction of computer-based decision support. Decision support has to be based on GPs' real needs. The knowledge base of SWDESTAR combines representations based on the Arden Syntax (for reminders) and a semantic net containing the medical factual knowledge. The representations facilitate knowledge sharing with other systems.

The comparisons show that several of the HIOS+ approaches also were applied in the four DSSs. The approaches concern the integration of passive and active decision support with a GP information system, the use of the Arden Syntax, the use of models or parts of models, the use of classifications, and the high involvement of GPs in the development process.

## Conclusions

With PDM, users' needs were determined. Functional and technical specifications were obtained through evolutionary prototyping. HIOS was installed in GP practices. It was positively evaluated by GPs and researchers. Requirements formulated in the checklists at the start of the HIOS+ development were successfully implemented. A blackboard system approach and MLMs were used. HIOS+ was tightly coupled with HIOS. Some decision support modules were evaluated, all with positive results regarding e.g. the assistance given in decision making and the prevention of misdiagnoses. All aims and objectives of HIOS and HIOS+ were reached. HIOS and HIOS+ both can be used to screen and monitor patient populations and offer diagnostic and disease management assistance.

## References

1. Allan D, Quinlan C. Making sense of computers in general practice. Oxford: Radcliffe Medical Press, 1995.
2. Rodnick JE. Evaluating feasibility and selection of computers in family medicine. *Journal of Family Practice* 1984;19:86-92.
3. Cruickshank PJ. Computers in medicine: patients' attitudes. *Journal of the Royal College of General Practitioners* 1984;34:77-80.

4. Curtis P. Computers in British general practice. *Journal of Family Practice* 1984;98-103.
5. Gilliland AEW, Mills KA, Steele K. General practitioner records on computer -- Handle with care. *Family Practice* 1992;9(4):441-450.
6. Ramsey CN, Grant WD. Administrative applications of computers in family practice. *Journal of Family Practice* 1984;19:80-85.
7. Schmittling G. Computer use by family physicians in the United States. *Journal of Family Practice* 1984;19:93-97.
8. Timpka T, Nyce JM. Dilemmas at a primary health care center: a baseline study for computer-supported cooperative health care work. *Methods of Information in Medicine* 1992;31(3):204-209.
9. Weiss B. Computer applications in family medicine: Too little, too late? *Family Medicine* 1994;26(7):428-429.
10. Williams RB, Boles M, Johnson RE. Patient use of a computer for prevention in primary care practice. *Patient Education Counseling* 1995;25(3):283-292.
11. Williams RL, Flocke SA, Zyzanski SJ, Mettee TM, Martin KB. A practical tool for community-oriented primary care community diagnosis using a personal computer. *Family Medicine* 1995;27(1):39-43.
12. The Adam Project Team, Deliverable 1.2: A general practitioners reference model - global architecture definition, AIM project 1023. Brussels: European Community, 1990.
13. Dupuits FMHM, Hasman A. Designing HIOS+. Proceedings of the IMIA Working Conference on Software Engineering in Medical Informatics SEMI 1990. Amsterdam: North-Holland Elsevier Science Publishers BV, 1991:181-193.
14. Simon HA. A behavioral model of rational choice. *Quarterly Journal of Economics* 1955;69:99-118.
15. American Society for Testing and Materials, E 1460 Standard Specification for Defining And Sharing Modular Health Knowledge Bases (Arden Syntax for Medical Logic Modules). Philadelphia: ASTM, 1992:539-587.
16. Hripsak G, Clayton PD, Pryor A, Haug P, Wigertz OB, van der Lei J. The Arden Syntax for Medical Logic Modules. In: Miller RA, ed. Proceedings of the Annual Symposium on Computer Applications in Medical Care 1990. Amsterdam: Elsevier Publishers, 1990:200-204.
17. Hripsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden Syntax. *Computers in Biomedical Research* 1994;27:291-324.
18. McDonald CJ, Hripsak GH. Data exchange standards for computer-based patient records, in Aspects of the Computer-based Patient Record. In: Ball MJ, Collen MF, eds. New York: Springer-Verlag, 1992:157-164.
19. Aho AV, Ullman JD. Principles of Compiler Design. Reading: Addison-Wesley Publishing Company, 1979.
20. Bailey JE. Development of an instrument for the management of computer user attitudes in hospitals. *Methods of Information in Medicine* 1990;29(1):51-56.
21. Bailey JE, Pearson SW. Development of a tool for measuring and analyzing computer user satisfaction. *Management Science* 1983;29(5):530-545.
22. Dupuits FMHM, Hasman A. User satisfaction of general practitioners with HIOS+, a medical decision support system. *Computer Methods and Programs in Biomedicine* 1995;47:183-188.
23. Weed LL. Knowledge coupling - New premises and new tools for medical care and education. New York: Springer Verlag, 1991.
24. Hofmans-Okkes IM, Lamberts H. The International Classification of Primary Care (ICPC): New applications in research and computer-based patient records in family practice. *Journal of Family Practice* 1996;13(3):294-302.
25. Klinkman MS, Green LA. Using ICPC in a computer-based primary care information system. *Family Medicine* 1995;27(7):449-456.
26. Lamberts H, Wood M. ICPC - International Classification of Primary Care, Prepared for the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA) by the ICPC Working Party. Oxford: Oxford University



- Press - Oxford Medical Publications, 1987.
27. WONCA, ICHPPC-II-Defined - International Classification of Health Problems in Primary Care. Oxford: Oxford University Press, 1983.
28. Leclercq RMFM, de Geus CA. Registration network family practices: how to classify physical examination. In: WONCA committee eds. Abstracts of the 11th Conference of the World Organization of National Colleges, Academics and Academic Associations of General Practitioners/Family Physicians 1986. Glasgow: Bell and Bain Ltd, 1986:303.
29. Kristensen FB, Kelstrup J, Kohlbau C, Lassen LC. Computer-based longitudinal recording of episodes of care in general practice using the International Classification of Primary Care (ICPC). *Scandinavian Journal of Primary Health Care* 1993;11(1):53.
30. Mulloy JV, Leuschen M, Rowe BH. Computer-based patient encounter tracking - Development of a system for family medicine residents. *Canadian Family Physician* 1995;41(10):1742-1753.
31. Barrows RC, Johnson SB. A data model that captures clinical reasoning about patient problems. *Proceedings of the Annual Symposium on Computer Applications in Medical Care*. Washington, 1995:402-5.
32. Friedrich G, Gottlob G, Nejd W. Generating efficient diagnostic procedures from model-based knowledge using logic programming techniques. *Computers in Mathematical Applications* 1990;20(9/10):57-72.
33. Keravnou ET, Johnson L. Towards a generalized model of diagnostic behaviour. *Knowledge-Based Systems* 1989;2(3):165-177.
34. Dupuits FMHM, Hasman A, Ulrichs EMJJ. Modelling and knowledge (re)presentation within HIOS+, in *AIME 91 - Proceedings of the Third Conference on Artificial Intelligence in Medicine*, eds. Stefanelli M, Hasman A, Fieschi M, Talmon J. Berlin/Heidelberg: Springer-Verlag, 1991:95-104.
35. Dupuits FMHM. The use of the Arden Syntax for MLMs in HIOS+, a decision support system for general practitioners in The Netherlands. *Computers in Biology and Medicine* 1994;24(5):405-410.
36. Dupuits FMHM, Hasman A. Evaluation of a decision support system in a medical environment. *International Journal of Bio-Medical Computing* 1995;38:155-165.
37. Dupuits FMHM. Evaluation of GPs' medical knowledge by a decision support system. *Computer Methods and Programs in Biomedicine* 1995;48(1,2):91-96.
38. Hripcsak G, Allen B, Cimino JJ, Lee R. Access to data: comparing AccessMed with Query by Review. *Journal of the American Medical Informatics Association* 1996;3(4):288-299.
39. Hartigan JA. *Bayes' theory* New York: Springer Verlag, 1983.
40. Brooks GJ, Ashton RE, Pethybridge RJ. DERMIS: a computer system for assisting primary-care physicians with dermatological diagnosis. *British Journal of Dermatology* 1992;127(6):614-619.
41. Jousimaa J, Kunnamo I. PDRD -- a computer-based primary care decision support system. *Medical Informatics (London)* 1993;18(2):103-112.
42. Elson RB, Connelly DP. Computerized decision support systems in primary care. *Primary Care* 1995;22(2):365-384.
43. Linnarsson R. Decision support for drug prescription integrated with computer-based patient records in primary care. *Medical Informatics (London)* 1993;18(2):131-142.

## **Chapter 4**

### **Patients, methods, and techniques in the patient- oriented study of thyroid disorder prediction instruments**

The references <sup>7,25,26,28-30</sup> included in the List of Publications also concern the subjects dealt with in this chapter.

## Introduction

This chapter describes the methods and techniques used in data collection and data analysis during and the patient population of the patient-oriented study of thyroid disorder prediction instruments. The aim of the prospective study was to investigate the capabilities of thyroid disorder prediction instruments in predicting the adequacy of Thyroid-Stimulating Hormone (TSH) test requests on the basis of patient data. A prospective study was needed because a retrospective review of laboratory test request forms had been unsuccessful due to the fact that General Practitioners (GPs) had registered too few signs/symptoms (clinical data) on these forms<sup>1</sup>.

Patient data were acquired at general practices from 1 September 1993 until 1 July 1994. The GPs at these general practices were affiliated within 'the "Heuvelland" network of GPs', which encompasses 87 GPs in 61 general practices in the Maastricht region. Network GPs can request diagnostic tests to be performed at the University Hospital Maastricht via the hospital's Diagnostic Coordinating Centre (DCC). The DCC provides feedback to the GPs regarding their request behaviour using guidelines created with and accepted by the GPs. This feedback has positive results in that it leads to a quantitative reduction and a qualitative improvement of requests. The DCC's aim is to stimulate an as high as possible rationality in the test request behaviour of GPs regarding diagnostic tests.

## Methods/techniques used in data acquisition

Use was made of general procedures, study-specific procedures, questionnaires and the DCC document.

### *General procedures*

#### The GP request procedure

In accordance with the general GP request procedure still in use, a GP either specifically or nonspecifically (in combination with other tests) requested a TSH IRMA test for each patient identified at a participating general practice during the data acquisition period by completing a laboratory test form. On this form, the GP provided the hypothesis for request (which was difficult to decide on in some cases) together with clinical patient data, the request reason, and the laboratory tests. The form was taken by the patient to the hospital, where it was processed by the DCC staff. The DCC coordinated requests at the hospital's end and served as a GP/hospital intermediary.

#### The thyroid laboratory test procedure

In accordance with the general thyroid laboratory test procedure, a blood sample

of a patient was taken either directly at the hospital (if the general practice was situated inside the city) or by visiting hospital staff at the general practice (if the general practice was situated outside of Maastricht) to determine the TSH concentration. The sample was analysed at the hospital's Clinical Chemistry Laboratory. If a TSH test result was abnormal ( $\leq 0.2$  mU/l or  $\geq 5.0$  mU/l), a further conclusive Free Thyroxine (FT<sub>4</sub>) test (abnormal if  $< 10$  pmol/l or  $\geq 24$  pmol/l) was performed<sup>2</sup> automatically to determine the diagnosis of clinical/subclinical hyperthyroidism or clinical/subclinical hypothyroidism. The laboratory staff notified the DCC staff of the test results, who in turn informed the GP. TSH and FT<sub>4</sub> test results were interpreted according to the rules in table 1.

**Table 1.** Interpretation of TSH and FT<sub>4</sub> test results

CAT. \ TSH+FT <sub>4</sub>	TSH	Operator	FT <sub>4</sub>
Clin. hyperth.	$\leq 0.2$ mU/l	AND	$\geq 24$ pmol/l
Subcl. hyperth.	$\leq 0.2$ mU/l	AND	$\geq 10$ AND $< 24$ pmol/l
Clin. hypoth.	$\geq 5.0$ mU/l	AND	$< 10$ pmol/l
Subcl. hypoth.	$\geq 5.0$ mU/l	AND	$\geq 10$ AND $< 24$ pmol/l
Euthyroid.	$> 0.2$ mU/l AND $< 5.0$ mU/l		

The range mentioned for an abnormal FT<sub>4</sub> was valid during the entire 10-month data acquisition period. Later, however, this range was adjusted by the laboratory staff to  $< 8$  pmol/l or  $\geq 18$  pmol/l because since then the more sensitive FT<sub>4</sub> of Wallac (Two-Step Method) test is used. The ranges were subsequently applied in the ongoing prospective study.

### *Study-specific procedures*

#### Inclusion criteria

Patients were included in the study if they were aged 15 years or over, were not pregnant, and used no thyrostatica (inhibition) or thyromimetics (supplementation) or had been using such drugs for 6 months or longer.

#### The questionnaire procedure

For every included patient for whom a TSH test was requested, a GP answered a questionnaire about diagnostic data which upon completion either was mailed directly to the hospital in a prepaid answering envelope or was given to the patient to be handed to a

DCC employee at the hospital. A maximum time period of 7 days for obtaining the questionnaires was imposed in order to avoid a bias in a GP's thinking due to the fact that he/she was informed about a TSH test outcome prior to answering the questionnaire. A GP also handed a patient questionnaire to every patient for whom a TSH test was requested. A patient sent a completed questionnaire to the hospital in a prepaid answering envelope, gave it to the hospital nurse taking the blood sample at the general practice to be handed to a DCC employee, or passed it directly to a DCC desk receptionist at the hospital. The reminder procedure was identical for GPs and patients in that they were reminded either over the telephone or in writing of having to return their questionnaires, and as a rule was initiated 3 weekdays after a TSH test request filing date. At the DCC desk, a participating patient additionally always was asked if he/she had already obtained a questionnaire (the safety net). During the data acquisition period, a patient questionnaire was issued by a DCC employee in 73% and by a GP in 27% of the instances.

#### *Questionnaires and the DCC document*

##### Questionnaire development

Diagnostic data indicative of thyroid disorders were acquired using the GP and patient questionnaires. These questionnaires predominantly contained dichotomous questions ('yes'/'no' answers) and some multiple-choice questions concerning signs/symptoms (items) of thyroid disorders. Several experts, including an endocrinologist and statisticians, were consulted in the creation of both questionnaires. Items used had been mentioned in textbooks and indices<sup>3-7</sup> or were suggested by the experts. Prior to the prospective study, no thyroid indices existed for the GP environment. Existing indices, which are described in chapter 2, were originally designed for diagnostic use in a hospital environment.

##### The patient questionnaire

The patient questionnaire (Appendix II) contained 54 questions about diagnostic items and additional aspects which might effect thyroid gland functioning like smoking habits<sup>8-10</sup>, alcohol consumption<sup>11-13</sup>, and menstruation<sup>14-16</sup>. The questions were phrased in such a way that laymen could be expected to answer them reliably.

##### The GP questionnaire

The GP questionnaire (Appendix III) contained 7 questions about the thyroid disorder hypothesis, physical examination results, past thyroid laboratory test results, the patient's thyroid disease history, and the patient's medication. It was deliberately kept brief in order to guarantee GPs' involvement and to assure that the data provided by the

GPs would be reliable. Aspects taken into account were that questionnaire completion should not take too much of a GP's already limited time and that the questionnaire only should ask about items which GPs are used and experienced to assess (not the Achilles' tendon reflex, e.g.).

#### Questionnaire validation

The quality of the questionnaires was expressed in terms of the internal consistency which can be statistically represented in Crohnbach's alphas and is a measure of how strongly questionnaire items are related to each other<sup>17</sup>. Related items concerned (i) nervousness, agitation, and irritability; (ii) forgetfulness and a decreased concentration ability; and (iii) preference for heat, dry skin and dry hair. Computed Crohnbach's alphas for these clusters were satisfactory and came to 0.79, 0.65, and 0.82 respectively.

The questionnaires, furthermore, were validated using a population of 221 patients visiting a health centre. The patients answered the patient questionnaire in the presence of a GP. The GP checked the answers. It was concluded that all questions had been correctly interpreted by 217 patients (98.2%). No changes were made in the questionnaires.

#### The DCC document

For every patient for whom a TSH test was requested by a GP, the DCC staff completed a DCC document (Appendix IV). The document was used to carry out administrative and managerial (reminder and feedback) functions regarding the study.

### **Patients**

Only patients who met the inclusion criteria were asked to participate in the study. The entire patient population was classified into groups of new, follow-up, and nonparticipating patients. The group of new patients was the patient population on which the development of the prediction instruments was based. The new patients, therefore, were further categorized into thyroid disease categories.

#### *The patient population*

During the data acquisition period, GPs requested TSH tests for 1,064 patients. Of these, 114 patients (11%) did not participate: 55 patients (5%) were able to participate but did not want to; 32 patients (3%) were not able to participate due to old age (11 patients; 1%), mental disorders/personal problems (15 patients; 1%), or language problems (6 patients; 1%); and 27 patients (3%) did not participate because their GPs had indicated in the first two months of the study not to participate after all. This was the case with 7 of

the 87 core network GPs (8%) in 6 general practices (10%). Their main reason for not participating was that they were already engaged in other research. The participating 950 patients (89%; 814 female and 136 male) comprised 668 (70%) new and 282 (30%) follow-up patients, which respectively constituted 63% and 26% of the entire patient population.

The request behaviour of GPs did not change during the data acquisition period. With the Wilcoxon Signed-Rank test, no significant difference ( $Z = -0.41$ ,  $p = 0.68$ ) was observed in this behaviour when comparing the first five months of the acquisition period with the last five months. The nonparticipation rate of patients over both periods, however, did change significantly in that a reduction in the number of nonparticipating patients was witnessed ( $Z = -2.63$ ,  $p = 0.009$ ). The reduction might have been a result of GPs more often familiarizing patients with the idea of being asked to complete a patient questionnaire and informing them about the aims of the prospective study.

### *Patient distributions*

#### The categorization of new patients

Table 2 presents the results of the categorization of the participating group of new patients ( $n = 668$ ) and the nonparticipating group of new patients ( $n = 84$ ) into the various thyroid disease categories. The latter group of new patients proved to be a random selection of the patient population with respect to aspects like disease category (table 2); age (24% of the patients were older and 76% younger than 60); and reasons for nonparticipation: 39 patients (46%) were unwilling to participate; 24 patients (29%) could not participate due to old age (8 patients; 10%), mental disorders/personal problems (11 patients; 13%), or language problems (5 patients; 6%); and 21 patients (25%) did not participate due to the nonparticipation of their GPs.

**Table 2. The categorization of new patients**

CAT. \ PAT.	Part. new patients		Nonpart. new pat.	
Clin. hyperth.	37	5.5%	5	6.0%
Subcl. hyperth.	21	3.1%	4	4.8%
Clin. hypoth.	8	1.2%	1	1.2%
Subcl. hypoth.	37	5.5%	7	8.3%
Euthyroid.	565	84.6%	67	79.7%
<b>TOTAL</b>	<b>668 = 63%</b>		<b>84 = 8%</b>	

Age of the participating new patients in relation to gender

Table 3 presents the distribution of the participating new patients regarding age in relation to gender. The overview shows that the group of 668 patients comprised 112 males (17%) and 556 females (83%).

**Table 3. Age/gender overview of the participating new patients**

AGE \ GENDER	Male	Female
15 - 20	3	24
21 - 30	10	101
31 - 40	26	121
41 - 50	28	123
51 - 60	16	78
61 - 70	12	55
71 - 80	12	35
81 - 90	3	18
91 - 100	2	1
TOTAL	112	556

**Methods/techniques used in data analysis**

The quality of acquired data, computer-based data entry, and the data base were analysed. Items indicative of thyroid disease were also determined in order to create the prediction instruments.

*Data quality*

The quality of acquired data was monitored concerning completeness and correctness<sup>18</sup>. If an answer was incomplete or contradictory, a GP or patient was contacted by telephone and asked an explanation.

Data completeness

Most GP questionnaires were filled in completely. The patients' physical examination results, however, were an exception: only in 70% of the 668 cases, all physical examination results were given; in 19%, only some results were ascertained; and in 11%, a physical examination was not performed at all or results were not given.

Patient questionnaires were filled in completely and conscientiously in all cases. Occasionally, patients added explanatory remarks to answers or even glued entire



medication packages or drug prescriptions to the questionnaire.

#### Data correctness

Data correctness could not be determined conclusively. Sometimes, for instance, data (e.g. a disease) were mentioned, but related data (e.g. the medication for the disease's treatment) were not. This did not necessarily mean, however, that the provided data were incorrect (there are, e.g., diseases which do not necessarily need drug treatment). An indication of data correctness was obtained with the help of an unaffiliated panel of two GPs who independently compared answers to similar questions in the GP and the patient questionnaires. The panel indicated that 89.7% of the data matched ( $n = 668$ ); in 6.1% a GP had been inexhaustive; in 2.7% a GP and patient gave different answers; and in 1.5% a patient had been inexhaustive. The kappa value of the panel GPs was 0.95, which is an excellent degree of consensus<sup>19</sup>. The two panel GPs concluded that both GPs and patients answered the related questions consistently.

During questionnaire completion, GPs were not biased because laboratory test results were reported to them only after their questionnaires had been received (as was agreed upon at the start of the study). The skewed distribution of the response time (the time period between a TSH test request filing date and a questionnaire receival date) of GPs had a median of 19 days. Furthermore, 38% of the GP questionnaires were received within 7 days, 30% within 3 days, and 16% immediately, i.e. on the TSH test request filing date. The average response time of patients was within 5 days.

#### *Computer-based data entry and data base quality*

The quality of data entry and the data base was monitored with the in-house registration system THYPRED (THYroid PREDiction information system). THYPRED offered data entry facilities, performed data entry checks, made reports concerning e.g. reminders, and performed analytical functions in order to create demographic overviews.

All patient data were stored in the data base using THYPRED. The software immediately gave a warning when, e.g., an incomplete or inconsistent answer was entered. If a patient indicated to have experienced a change in voice lately (patient question 32), e.g., the nature of this change needed to be entered too. The data then were entered into the data base. All data in the data base were also manually compared with data on the questionnaires. The result was that no inconsistencies were found in the data base.

*Data analysis: from data to prediction instruments*

Data analyses performed prior to instrument creation

With SPSS, statistical analyses were performed on the acquired data of the 668 new patients. The data were analysed univariately, bivariately, and multivariately. Raw Odds Ratios (ORs) for each individual item were obtained in the univariate analysis. In the bivariate analysis, correlations between two items were determined to assess possible relationships. The multivariate analysis included a factor analysis through which clusters of independent items were determined. The factor analysis resulted in the three data clusters described earlier. These were used together with independent items in stepwise logistic regression.

Statistical techniques considered for instrument creation

A variety of multivariate statistical techniques, which can be used to predict a dichotomously dependent item from a set of independent items, were considered for instrument creation. Among these were multiple (least squares) regression analysis and (linear) discriminant analysis. Least squares regression poses difficulties when the dependent item can have two outcomes only, like diseased or healthy. In this case, assumptions necessary for hypothesis testing in multiple regression analysis are violated, e.g. the assumption of normally distributed errors. Another difficulty with multiple regression analysis is that predicted outcomes are not probabilities (not constrained to the interval between 0 and 1). Linear discriminant analysis does allow a direct prediction of group membership, but the assumption of multivariate normality of the independent items in the two groups is required for the prediction rule to be optimal. A better multivariate technique for estimating the probability of a dependent dichotomous item is logistic regression. This technique requires fewer assumptions than discriminant analysis. Even when the assumptions required for discriminant analysis are not satisfied, logistic regression still performs well.

Data analysis during instrument creation

On the basis of the previous considerations, stepwise logistic regression analysis was performed as part of the multivariate analysis with SPSS in order to determine distinctive items for inclusion in the prediction instruments for hyperthyroidism and hypothyroidism.

Stepwise logistic regression generally implies that, if there are many independent items in a logistic regression equation, the probability of an event is determined with the formula:

$$Probability(event) = \frac{1}{1 + e^{-S}}$$

where  $S$  is the linear combination  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N$ . In this formula, the  $X_i$  are items with numeric values and the  $\beta_i$  are regression coefficients which quantify the items' contributions to the probability. The justification for this formula is that the log of the odds, a number between  $-\infty$  and  $\infty$ , is a linear function. Given this general equation, items can be selected stepwise and corresponding coefficients computed. In the creation of a logistic regression equation, the likelihood ratio test is used to add items to the equation if  $p < 0.05$ . Whenever an item is added, the set of items selected so far is reanalysed. If an item has a  $p > 0.05$ , it is removed from the set of selected items and stepwise logistic regression is continued with the remaining items to be added to the equation. This stepwise process is ended when the probability of all remaining items is nonsignificant. For items in the equation, adjusted ORs and their 95% Confidence Intervals (CIs) are computed with  $e^{(\beta_i - 1.96 \times SE)} \leq OR \leq e^{(\beta_i + 1.96 \times SE)}$  in which  $SE$  is the Standard Error of  $\beta_i$ , and  $e^{\beta_i}$  is the OR. Logistic regression has proven to be very robust in medical domains and an effective way of estimating probabilities of dichotomous items<sup>20-22</sup>.

The entire population is used to create the prediction instruments, which will be applied by GPs after they have decided to request a TSH test. For each of the instruments, pretest probabilities were ascertained along with the corresponding sensitivities, specificities, positive Predictive Values (PVs+), negative Predictive Values (PVs-), numbers of false positives, numbers of false negatives, positive Likelihood Ratios (LRs+), and negative Likelihood Ratios (LRs-). Based on this information, Receiver Operating Characteristic (ROC) curves<sup>23</sup> were created, which are graphical representations of the usefulness of an instrument by setting sensitivity against 1-specificity.

#### Instrument evaluation

An evaluation of the instruments was performed in accordance with the Grouped Jackknife Method<sup>24,25</sup>. The entire population was divided into two groups: one large group for instrument creation (using those items which are selected for inclusion in the instruments created with the entire population) and one small group (about 100 persons) for instrument evaluation. This partitioning was done in such a way that all patients were once evaluated during an iterative phase<sup>26</sup>. In all iterative phases, different learning/test sets and all available items were used. However, none of the prediction instruments created in the iterative phases will be used in practice. Sensitivities, specificities, PVs+, PVs- were determined by adding all iterative phase results. CIs for sensitivity and

specificity were determined either by using a table<sup>27</sup> if  $n < 100$  or by using the formula  $p \pm 1.96 \times (\sqrt{(p \times (1-p)/n)})$  if  $n \geq 100$ .

## References

1. Dupuits FMHM. The Quality of GP Referral Data. In: Teasdale S, Bradley P, eds. Proceedings of the Annual Conference of the Primary Health Care Specialist Group of the British Computer Society 1993. Worcester: PHCSG, 1993:88-102.
2. Massey KL. The Laboratory Approach to Thyroid Disease: A Stepped Approach. *Can Fam Physician* February 1988;34:365-368.
3. Crooks J, Murray IPC, Wayne EJ. Statistical Methods Applied to the Clinical Diagnosis of Thyrotoxicosis. *Quart J Med* April 1959;XXVIII(110):211-234.
4. Wayne EJ. Clinical and Metabolic Studies in Thyroid Disease (The first of two Lumleian Lectures delivered before the Royal College of Physicians of London on April 14 and 16, 1959). *Br Med J* 1960;2 January:1-11.
5. Gurney C, Owen SG, Hall R, Roth M, Harper M, Smart GA. Newcastle Thyrotoxicosis Index. *Lancet* 1970;19 December:1275-1278.
6. Wayne EJ. Clinical and Metabolic Studies in Thyroid Disease (The second of two Lumleian Lectures delivered before the Royal College of Physicians of London on April 14 and 16, 1959). *Br Med J* 1960;9 January:78-90.
7. Billewicz WZ, Chapman RS, Crooks J, Day ME, Gossage J, Wayne E, Young JA. Statistical Methods Applied to the Diagnosis of Hypothyroidism. *Quart J Med* April 1969;XXXVIII(150):255-266.
8. Müller B, Zulewski H, Kunz M, Mitrache Ch, Greber P, Staub JJ. Influence of Cigarette Smoking on the Severity of Hypothyroidism. *Eur J Clin Invest* 1993;23(60-Suppl.1):A12.
9. Petersen K, Lindstedt G, Lundberg P-A, Bengtsson C, Lapidus L, Nyström E. Thyroid Disease in Middle-aged and Elderly Swedish Women: Thyroid-related Hormones, Thyroid Dysfunction and Goitre in Relation to Age and Smoking. *J Intern Med* 1991;229:407-414.
10. Ericsson U-B, Lindgärde F. Effects of Cigarette Smoking on Thyroid Function and the Prevalence of Goitre, Thyrotoxicosis and Autoimmune Thyroiditis. *J Intern Med* 1991;229:67-71.
11. Tran PD, Leclerc A, Chastang JF, Goldberg M. Prevalence of Health Problems in the Gazel (EDF-GDF) Cohort: Regional Distribution and Disparities. *Revue d'épidémiologie et de santé publique* 1994;42(4):285-300.
12. Orrego H, Blake JE, Blendis LM, Compton KV, Israel Y. Long-term Treatment of Alcoholic Liver Disease with Propylthiouracil - Part 2: Influence of Drop-out Rates and of Continued Alcohol Consumption in a Clinical Trial. *J Hepatol* 1994;20(3):343-349.
13. Hegedus L. Thyroid Size Determined by Ultrasound - Influence of Physiological Factors and Non-thyroidal Disease. *Dan Med Bull* 1990;37(3):249-263.
14. Knoepfel MC, Landolt AM, Froesch ER. TSH-produzierende Hypophysenadenome und sekundäre Hyperthyreose (Fallvorstellung und Literaturübersicht). *Schweiz Med Woch* 1989;119(34):1159-1163.
15. Tamai H, Mizuno O, Takaki A, Kiyohara K, Komaki G, Matsubayashi S, Kuma K, Kumagai LF, Nagataki S. Heterogeneity of Serum Prolactin in Patients with Menstrual Disorder in Conjunction with Hyperthyroxinemia. *J Endocrin Invest* 1991;14:679-684.
16. Blum M, Blum G. The Possible Relationship Between Menorrhagia and Occult Hypothyroidism in IUD-wearing Women. *Advan in Contracept* 1992;8:313-317.
17. Dawson-Saunders B, Trapp RG. *Basic & Clinical Biostatistics*. 2nd ed. London: Prentice-Hall International Inc., 1994.
18. Weed LL. *Knowledge Coupling - New Premises and New Tools for Medical Care and Education*. New York: Springer Verlag, 1991:163-166.
19. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: Wiley Publishers,

- 1981.
20. Barnas C, Fleischhacker WW, Whitworth AB, Schett P, Stuppach C, Hinterhuber H. Characteristics of Benzodiazepine Long-term Users: Investigation of Benzodiazepine Consumers among Pharmacy Customers. *Psychopharm Berl* 1991;103(2):233-239.
  21. Stevens MM, Freeman DH, Mott LA, Youells FE, Linsey SC. Smokeless Tobacco Use among Children: The New Hampshire Study. *Am J Prev Med* 1993;9(3):160-167.
  22. Sendrowski DP. Hyperthyroidism. *Optom Clin* 1994;3(4):87-97.
  23. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology - A Basic Science for Clinical Medicine*. 2nd ed. Boston: Little, Brown and Company, 1991.
  24. Miller RG. The Jackknife - A Review. *Biometrika* 1974;61:1-15.
  25. Schouten HJA. *Statistical Measurement of Interobserver Agreement* [Doctoral Thesis at the Erasmus University Rotterdam]. Utrecht: Drukkerij Elinkwijk B.V., 1985.
  26. Norusis MJ. *SPSS/PC+ Advanced Statistics 4.0 for the IBM PC/XT/AT and PS/2*. Chicago: SPSS Inc, 1990:B11.
  27. Lentner C, Diem K, Seldrup J. *Geigy Scientific Tables - Volume 2: Introduction to Statistics, Statistical Tables, Mathematical Formulae*. 8th, revised and enlarged ed. Basle: Ciba-Geigy Corporation, 1982.

## **Chapter 5**

### **Evaluation of a decision support system in a medical environment**

The article presented in this chapter was published in

Dupuits FMHM, Hasman A. Evaluation of a decision support system in a medical environment. International Journal of Bio-Medical Computing 1995;38:155-165.

The references <sup>3,18,23,24,35</sup> included in the List of Publications also concern the research described here.

## **ABSTRACT**

This paper describes the impact of a decision support system on the quality of recorded diagnoses and the completeness of medical records. The assumption in the study was that for quality assessment purposes enough appropriate data have to be recorded in an electronic medical record in order to be able to justify diagnostic decisions. The hypotheses were tested that active decision support will lead to a better quality of recorded diagnoses and more complete medical records. Three groups of ten GPs were presented 10 cases each. The GPs had to enter data about these cases in a GP information system. One group of GPs was not supported. The second group was presented the ICHPPC-II-Defined criteria which had to be fulfilled for an entered diagnosis to be justified (passive support). In a third group, the GPs were asked those data which were needed to justify an entered diagnosis and which were not already registered by the GPs (active support). It could be shown that the last group of GPs entered diagnoses of a better quality than the other two groups. These two groups also did enter so little data that most of their diagnoses could not be justified. It was concluded from the study that active decision support will lead to better diagnoses and a more complete medical record which can be used for quality assessment.

## **KEYWORDS**

Decision Support System, Family Medicine, Quality Assurance, ICPC, Code Validation, Electronic Medical Record

## **Introduction**

Information systems are increasingly used by GPs. In The Netherlands, about 70% of GPs use an information system in daily work. Approximately half of this number of GPs do not only have an administrative module at their disposal but also a medical module. An estimated 10-15% of all Dutch GPs actually make use of an electronic medical record in which they store demographic and medical data of patients. In The Netherlands, medical data are registered in a problem-oriented way: use is made of the SOAP approach in which data are recorded as Subjective findings (S), Objective findings (O), Assessment of signs, symptoms, and diagnoses (A), and Plan (P). Moreover, the Dutch version of the International Classification of Primary Care (ICPC) is used for coding complaints, diagnoses, and therapies<sup>1</sup>.

The national 'ICPC Thesaurus Project' was initiated by the Dutch College of General Practitioners and several universities to promote a better (coded) registration of medical patient data by GPs by improving the accessibility of the Dutch ICPC version. The task of the Maastricht University (formerly called the University of Limburg) concerned the disclosure of free-text criteria belonging to ICPC codes for diagnoses and complaints. These criteria are described in the International Classification of Health Problems in Primary Care (ICHPPC)-II-Defined<sup>2</sup>. The task involved an investigation of the usefulness of decision support in the validation of ICPC codes which are registered by GPs during daily practice.

Decision support tools to be developed should be able to check the validity of a registered ICPC code in a patient's case using the ICHPPC-II-Defined criteria. Two approaches could be followed. In the first place, criteria belonging to a certain ICPC code could be displayed as soon as this code was entered. The GP would be able to decide whether he/she entered the correct code on the basis of the description (passive decision support). In the second place, available free-text criteria could be translated into computer-processable queries which could be used for querying a medical record in order to check the validity of an entered ICPC code automatically (active decision support). It was hypothesized that automated active decision support will lead to better coding when compared with passive decision support or no support at all. It also was hypothesized that only active decision support will lead to a better documented patient record. A more complete medical record implies that decisions made can be justified on the basis of data in the medical record.

HIOS+<sup>3-5</sup>, which was developed at the Department of Medical Informatics, is a medical registration system in which decision support facilities are integrated. In order to be able to test the usefulness of decision support in the validation of entered ICPC codes, two modules were added to HIOS+: the Free-text Formalizer (FF) and the Formal Interpreter (FI).

In this paper, the difficulties which were met when formalizing free-text criteria, the Free-text Formalizer and the Formal Interpreter, the design of the experiment, and the experiment results are described and discussed.

### **Formalization of criteria**

ICHPPC-II-Defined criteria are available for 356 of the 685 ICPC codes (of which 322 concern complaints and 363 concern diagnoses). Of these 356 codes, 129 refer to complaints and 227 to diagnoses.

Since the ICHPPC-II-Defined criteria are free-text criteria, they had to be translated (formalized) into queries which are interpretable by a computer. Use was made of the Arden Syntax for Medical Logic Modules (MLMs)<sup>6</sup> to express these criteria. In figure 1, two examples of the use of the Arden Syntax are presented. These examples show what conditions can be expressed with the Arden Syntax. It can be stated, for example, that one out of a number of symptoms or signs should be present, or that, for the diagnosis uncomplicated hypertension to be justified, among other things the diastolic blood pressure should be higher than 95 mmHg when measured twice in two separate weeks or a diastolic blood pressure higher than 120 mmHg should be measured twice.



**Figure 1.** Examples of the expressiveness of the Arden Syntax. The queries of two diagnoses are presented.

---

**GOUT:**

IF ANY 1.00 (severe acute inflammation joints, (anamnesis uric acid AND (kidney dysfunction OR kidney insufficiency)), tophus) AND ANY 1.00 (((GENDER EQ male) AND (blood uric acid GT 0.42)), ((GENDER EQ female) AND (blood uric acid GT 0.34)), (renal biopsy AND uric acid deposits), (uric acid AND tophus), (uric acid crystals AND joint aspirate)).

**UNCOMPLICATED HYPERTENSION:**

IF (AGE GT 18) AND ANY 1.00 (((systolic blood pressure GT 160) AND (diastolic blood pressure GT 95) 2 TIMES) AFTER 1 WEEK), ((diastolic blood pressure GT 120) 2 TIMES)) AND (NOT heart problems) AND (NOT kidney problems) AND (NOT brain problems).

---

The biggest problem met in criteria formalization was the vagueness of the free-text descriptions of certain criteria. The description 'characteristic histological picture', for example, had to be made explicit by specifying the elements which compose such a description. Since it was not possible to elucidate all vague criteria descriptions due to the limited duration of the project, only the more or less clear criteria belonging to 101 ICPC codes were formalized.

During the process of formalization, some additional problems were encountered. Firstly, the results of tests are valid for only a certain period of time depending on the test. If test results are mentioned in a query, their "age" should be considered. For those test results for which no expiry period was mentioned in the criteria, a period of one month was used. If expiry periods were mentioned in the criteria, these were included in the query. The example of uncomplicated hypertension in figure 1 shows how these expiry periods can be taken into account. Secondly, a condition, like macrocytic anaemia, which is included in the criterion of a diagnosis can itself be established by another criterion. This criterion is then embedded in the criterion of the diagnosis, replacing the original condition.

Another formalization problem concerned the coding of some item categories. Various item categories are used in the criteria: complaints, diagnoses, physical examination results, age, habits like smoking or alcohol consumption, work environment, laboratory test results, etc. The ICPC does not provide codes for each of these categories. Therefore, new codes were created. These codes were added to the ICPC for in-house use only. On the SOAP

level, added codes start with a 'V' to distinguish them from ICPC codes.

An independent national GP panel evaluated the 101 created queries. In total, 70 queries were evaluated positively; 24 showed minor and 7 major imperfections. These defects were corrected.

For transferability reasons, the queries were encoded into MLMs using the Arden Syntax. A MLM is a structured, independent module containing medical knowledge. It consists of a maintenance section (containing information about the author of the MLM in question), a library section (containing information about possible links to related medical subjects) and a knowledge section (containing the query among other things).

### **Query creation and query processing**

The way queries are generated and the way in which they are used will now be explained. Both the Free-text Formalizer (FF) and the Formal Interpreter (FI) will be discussed.

#### *The Free-text Formalizer*

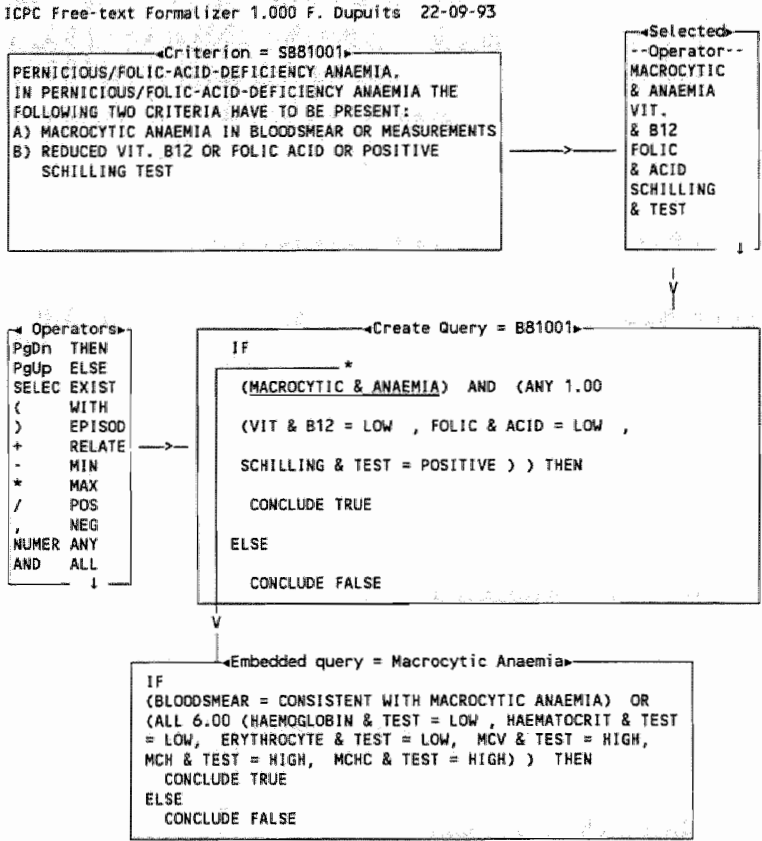
The FF module is a tool with which (medical) experts can create queries by formalizing free-text criteria. When an expert has activated FF, the system presents a menu with the following options: Display, Select, Link, Create, Evoke, Store, Edit, and Quit. In the display mode, FF presents an ICHPPC-II-Defined criterion (in free text) to the expert. An example of such a criterion is the one belonging to ICPC code B81001 representing pernicious/folic-acid-deficiency anaemia:

*In pernicious/folic-acid-deficiency anaemia, the following two criteria have to be present:*

- a. Macrocytic anaemia in bloodsmear or measurements
- b. Reduced vitamin B<sub>12</sub> or folic acid or positive Schilling test

After having activated the select option, the expert identifies relevant topics (keywords) in the free text of the criterion. In the example, 'macrocytic anaemia', 'bloodsmear', 'measurements', 'reduced vitamin B<sub>12</sub>', 'folic acid', and 'positive Schilling test' can be chosen as topics. If a topic consists of more than one term, the expert indicates this fact to FF and selects the individual terms which are then automatically concatenated by FF using the '&' sign ('macrocytic & anaemia', for example, is the translation of the compound topic 'macrocytic anaemia').

**Figure 2.** Query creation. In the windows, the free-text criteria, the selected topics, some of the operators, and the query are shown. The embedded query on macrocytic anaemia in the query on pernicious/folic-acid-deficiency anaemia is also shown. The window headed 'Selected' enables jumping to the window headed 'Operators' by selecting the '--operator--' option.



In the link mode, FF searches its ICPC data base files for one or more ICPC codes whose descriptions (partially) match a given topic. The expert chooses one or more of these codes and links them to the topic. If FF does not come up with a matching ICPC code, the expert can add a code by using the 'insert' key. In that case, a V code is generated. The code VB81007 is an example of a code added by the expert and therefore is not a regular ICPC code; 'B81' indicates that this code concerns the criterion belonging to ICPC code B81; and '007' is a sequence number generated automatically by FF (several new codes can be generated for an individual criterion). The user specifies the description of the added code;

the description of the V code presented above, for example, is 'blood MCV test'. This topic can now also be used in other queries. The code only is a reminder of the criterion in which the topic was encountered first. Later, when the criterion is executed by FI, the system looks for the ICPC codes and/or in-house codes in the medical record.

After having activated the create option, the expert combines the topics into a query with the help of Arden Syntax operators. The topics are selected one by one and then combined with the help of one or more operators. Operators can be logical, for example 'ANY 1.00' and 'GT', or temporal like 'BEFORE' and 'AFTER'. In figure 2, windows containing the information needed to create a query on pernicious/folic-acid-deficiency anaemia from the free-text criterion are presented. In this figure, an example is shown of a topic ('macrocytic anaemia') which itself has a criterion. Later, in the processing of the primary query (B81001), the parser removes the topic and replaces it by its embedded query. Syntactical checks are performed at the end of the creation process.

Next to the criteria pertinent to the diagnoses, the ICHPPC-II-Defined criteria also define separate inclusion and exclusion criteria. In the evoke mode, the expert can link the created query with these inclusion and/or exclusion criteria.

With the store option, the MLM is stored. Stored MLMs can be changed at any time by activating the edit mode. With the quit option, the expert returns to a HIOS+ work screen.

### *The Formal Interpreter*

The FI module gives decision support to GPs by checking the validity of entered ICPC codes. When a GP enters an ICPC code, FI activates the corresponding query (if available). FI then checks whether the (coded) query items are present in the patient's record. The GP is questioned about missing items necessary for the evaluation of the query. These missing items can be added to the patient's record. In figure 3, FI asks the GP to enter the haematocrit reading for Mr. W.J.N. Example. In this case, the GP can enter whether the outcome was normal, low, or high. In the 'S' column, the GP can indicate whether the data entered in the current encounter were already known from a previous encounter ('1') or not ('2'). In the 'C' column, the GP can record his/her diagnostic confidence: a '1' for 'not sure' or a '2' for 'sure'. When processing, FI first processes the primary query (the query belonging to the ICPC code which is currently being checked) until the system meets an embedded query (a query which further specifies a condition included in the free-text criterion). After the embedded query has been processed, FI continues to process the primary query. If the outcome of the primary query is 'true', possible secondary queries (queries belonging to ICPC codes which are specified as inclusions (of which the processing outcomes

need to be 'true') or exclusions (of which the processing outcomes need to be 'false')) are evaluated. When a GP enters a wrong diagnosis, the query belonging to this diagnosis (if present) and not the query of the correct diagnosis is processed. Finally, FI states whether the entered ICPC code is valid. The GP can consult a log list to interpret FI's statement, and subsequently can have a look at the query. In case FI advised that the ICPC code was invalid, the GP still has the opportunity to store the chosen ICPC code.

**Figure 3. FI aks for a test result**

[illegible]

## Materials and methods

The impact of the FI program on the quality of entered diagnoses and the completeness of the medical record was evaluated. During a test, GPs entered medical data on the basis of true-to-life case descriptions which were drawn up by a GP prior to the test. The diagnoses of the cases were chosen on the basis of criteria availability and disease prevalence. An example of a case description used in the test is presented in figure 4. The correct diagnosis to be registered by the participating GPs on the basis of this case description was acute bronchitis.

Some details about the case descriptions are presented in figure 5 which contains the diagnostic ICPC code for each case description and the ICPC code description. Each GP was asked to diagnose the condition described in each case description and to enter all necessary data needed to justify the diagnosis using HIOS+. One additional case description was used to familiarize the GPs with HIOS+ and was not taken into account in the evaluation. The GPs were divided into three equally large groups which differed in the amount of decision

support given. Whereas group I did not get any support at all, group II was provided passive support by automatically presenting the criteria belonging to the entered ICPC codes. Group

**Figure 4. Example of a case description**

**CASE DESCRIPTION: COUGH**

Niels Boorman, born 30-11-1986, enters your consulting room, partly hidden behind his mother. He has a cough for almost a week now. At first the cough was especially active at night. However, for two days now, Niels also coughs a lot during daytime and expectorates sputum. He does not have a fever. On examination, you find that Niels breathes fastly. On auscultation, you hear dispersed coarse, moist and dry rhonchi.

Medical history:     1990 fever convulsions and influenza  
                          1990 little bead in nose  
                          1991 mumps

III was given active support: if there were insufficient appropriate data to justify a diagnosis, a GP was asked to enter more relevant data. The data entered by the participating GPs were evaluated by the GP who made the case descriptions, and, in case of incorrect diagnoses, also by an independent GP panel. When analysing the data, two phases were distinguished: phase 1 concerned the analysis of the data before a diagnosis was entered, and phase 2 concerned the analysis of the data after a diagnosis was entered and after the intervention of FI.

**Figure 5. Case description overview**

ICPC CODE	DESCRIPTION OF THE ICPC CODE
D12001	Constipation
U71001	Cystitis/oth urin. infec.
S88001	Contact dermatitis
S09001	Paronychia
R78001	Acute bronchitis
L77001	Sprains & strains ankle
F70021	Infect. conjunctivitis
X72032	Urogen. candidiasis
R80022	Influenza
K77001	Heart failure

## Results

In this part, the results of the experiment are presented. Results concerning the diagnostic correctness of the three groups, the change in confidence in the entered diagnostic codes after the intervention of FI, and the manner in which FI influences the completeness of the medical record are described.

**Table I. Misdiagnoses for each group**

Registered diagnosis:			Correct diagnosis:	
Group I				
C2	L16992	Ankle swollen	L77	Sprains and strains of ankle(s)
C3	R05001	Cough productive	R78	Bronchitis/bronchiolitis acute
C4	U70041	Pyelonephritis/pyelitis acute	U71	Cystitis/oth. urin. infect. non-venereal
C5	N01001	Headache	R80	Influenza proven without pneumonia
C6	K02992	Tightness of chest	K77	Cardiac decompensation
C6	R02031	Shortness of breath	K77	Cardiac decompensation
C7	F29991	Photophobia	F70	Infectious conjunctivitis viral
C9	S22993	Finger swollen	S09	Paronychia
C0	S92021	Dyshidrosis	S88	Contact dermatitis/eczema
C0	S92021	Dyshidrosis	S88	Contact dermatitis/eczema
C0	S92021	Dyshidrosis	S88	Contact dermatitis/eczema
Group II				
C1	A85001	Medication side effects	D12	Constipation
C2	L16994	Ankle swollen	L77	Sprains and strains of ankle(s)
C3	P81001	Pneumonia	R78	Bronchitis/bronchiolitis acute
C3	R29001	Other symptoms bronchi	R78	Bronchitis/bronchiolitis acute
C5	A03001	Fever	R80	Influenza proven without pneumonia
C7	F02001	Eye red	F70	Infectious conjunctivitis viral
C7	F03022	Flood of tears	F70	Infectious conjunctivitis viral
C9	S22994	Nail wall red	S09	Paronychia
Group III				
C1	A85001	Medication side effects	D12	Constipation
C8	A78041	Other infectious diseases	X72	Urogenital candidiasis proven
C0	S04051	Papules skin swollen	S88	Contact dermatitis/eczema

### Diagnostic correctness

The number of correct diagnoses was determined for each GP taking part in the test. Group I GPs misdiagnosed 11 of the 100 diagnoses, group II GPs 8 out of 100, and group III GPs 3 out of 100. In table I, the misdiagnoses and the corresponding correct diagnoses are presented for each group. The results of the analysis were verified by an independent panel of GPs.

In order to determine if statistically significant differences existed between the three groups, 'a non-parametric comparison of two groups - the Mann-Whitney test'<sup>7</sup> was applied. A statistically significant difference was found between group III and group I ( $P = 0.04$ ). The differences between group II and group III and between group I and group II were not statistically significant.

In group III, FI prevented users from misdiagnosing a case in six instances. ICPC code R81 was once and R95 twice changed into the correct diagnostic code R78. In addition, U70 was once corrected into U71, F71 into F70, and S87 into S88. In group II, one GP entered additional information after being shown the free-text criterion but did not change his incorrect diagnosis. In group I, no feedback was given and no additional information was added by the GPs after registering the diagnoses.

### *Diagnostic confidence*

In the test, group III GPs firstly could state their confidence in the diagnoses both before and after the intervention of FI. After the intervention of FI, GPs became more confident about their diagnoses in 18 instances. In one instance (which concerned an incorrect diagnosis), a GP became less confident. In the remaining 81 instances, the GPs' diagnostic confidence concerning registered diagnoses did not change.

Secondly, group III GPs could express whether they coded according to the criteria. In 27 instances, GPs initially indicated that they had not coded in accordance with the criteria. After using FI, they discovered that in these cases they actually had acted according to the criteria. In 7 instances, GPs indicated to have coded according to the criteria but later on found out that that was not the case. In the remaining 66 instances, confidence in coding in accordance with the criteria did not change and indeed the GPs coded according to the criteria.

### *Completeness of the medical record*

When using the GP information system as an electronic medical record, it is essential that those data are recorded which are needed to justify decisions. In the study, this was translated into the requirement that enough appropriate data should be registered in order for FI to be able to evaluate the query. It should be stressed that completeness does not mean that all possible data should be stored. This asks too much of a busy GP and only leads to information overload. A complete medical record is a medical record which contains all the necessary data to justify recorded diagnoses.

The analysis of registered data was divided into two parts. First, it was investigated



whether the query could be evaluated on the basis of the data entered in phase 1. The results are presented in table II.

Table II. Outcome of queries in phase 1

	Group I			Group II			Group III		
	Processable		Not processable	Processable		Not processable	Processable		Not processable
	True	False		True	False		True	False	
GP 1	0	0	9	1	0	8	1	0	9
GP 2	0	0	9	0	0	10	2	0	8
GP 3	1	0	8	0	0	10	1	0	9
GP 4	1	0	8	1	0	9	0	0	9
GP 5	0	0	8	2	0	6	1	0	9
GP 6	1	0	6	0	0	8	1	0	9
GP 7	2	0	8	2	0	8	1	0	9
GP 8	0	0	10	0	0	9	3	0	7
GP 9	0	0	8	0	0	9	1	0	8
GP 10	1	0	9	0	0	9	1	0	8

This table presents for every GP in every group how many queries could or could not be evaluated and how many of the "processable" queries resulted in a 'true' or 'false' outcome in those cases for which the GPs ultimately (at the end of registration) reached a correct diagnosis. For GP 1 in group II, for instance, 8 queries could not be processed because of incomplete data and for 1 "processable" query the outcome was 'true'. One case was diagnosed incorrectly and was therefore not taken into account here. Table II shows that the majority of queries could not be processed because not enough patient data were registered. In phase 1, no difference in completeness was expected between the groups since no feedback was yet given by FI. This was substantiated using the Mann-Whitney test. Indeed, no statistically significant differences could be determined between the three groups in phase 1.

A similar analysis of registered data was performed in phase 2. Results of this analysis are shown in table III. This table presents for every GP in every group how many queries could be processed and how many of these "processable" queries had a 'true' or 'false' outcome. Again only those cases for which the GPs ultimately (at the end of registration) reached a correct diagnosis were considered. Remember that in phase 2 the intervention of FI only might have changed the number of data stored in groups II and III. Table III shows that only in group III, because of FI's intervention, the majority of queries became "processable" with 'true' outcomes. A statistically significant difference between groups I and III ( $P = 0.0001$ ) and between groups II and III ( $P = 0.0001$ ) was established with the Mann-Whitney test.

**Table III. Outcome of queries in phase 2**

	Group I			Group II			Group III		
	Processable		Not processable	Processable		Not processable	Processable		Not processable
	True	False		True	False		True	False	
GP 1	0	0	9	1	0	8	10	0	0
GP 2	0	0	9	0	0	10	9	0	1
GP 3	1	0	8	0	0	10	10	0	0
GP 4	1	0	8	1	0	9	8	1	0
GP 5	0	0	8	2	0	6	8	0	2
GP 6	1	0	6	0	0	8	9	0	1
GP 7	2	0	8	2	0	8	10	0	0
GP 8	0	0	10	0	0	9	10	0	0
GP 9	0	0	8	0	0	9	8	1	0
GP 10	1	0	9	0	0	9	9	0	0

The queries which remained "unprocessable" and which concerned diagnoses registered by group III GPs were investigated in detail. It is possible that queries remain "unprocessable" because - although FI asks a GP for additional data in order to be able to process the query - a GP is not obliged to enter these data. In one instance, the item stating 'an increase in pain when straining the ankle' was not registered although it is necessary for diagnosing sprains and strains of the ankle(s). In two instances, the item stating 'the presence of a significant amount of bacteria' was not registered although it is necessary for diagnosing cystitis. Once, a positive result of a chlamydia test was not registered which is needed for diagnosing conjunctivitis.

Queries with the outcome 'false' concerning correct diagnoses registered by group III GPs also were investigated. Twice 'urethritis' was registered which in fact is an exclusion for cystitis (cystitis being the correct diagnosis).

The analysis of registered data also revealed that for those cases for which the GPs ultimately (at the end of registration) reached incorrect diagnoses, the queries were found "unprocessable" both in phase 1 and phase 2 with two exceptions. In group II, a GP registered an incorrect diagnosis for the influenza case although the query outcome was 'true'. In this case, the GP recorded the diagnosis at a too low level (at the level of symptoms). In group III, a GP registered an incorrect diagnosis for the constipation case although the query outcome was 'true'. This particular GP considered the patient's complaints to be due to side effects of the medication rather than due to constipation.

## Discussion

In this study, the impact of providing decision support to GPs when they enter diagnoses has been investigated. When using the GP information system as an electronic

medical record, the record contents should fulfil at least two requirements. In the first place, enough data should be stored in order to be able to determine how a GP arrives at certain decisions. In the study, diagnostic decisions were investigated. To this end, enough data should be available to justify a diagnosis. In the second place, the stored data should be appropriate in order to justify a diagnosis. ICHPPC-II-Defined criteria were used as golden standards to determine whether the stored data indeed were able to justify diagnostic conclusions. It was hypothesized that only with proper decision support GPs would enter enough appropriate data.

Although GPs were not forced to enter data even when support was provided, it appeared that in most cases in group III the necessary data were entered.

The analysis of registered data made clear that only presenting the criteria on the screen without explicitly asking a GP to enter the data does not lead to additional data entry. Apparently, a GP is usually confident about his/her diagnosis and does not consider further data entry worthwhile. Even in cases in which the diagnosis was incorrect, a GP did not change his/her diagnosis after being presented with the free-text criteria. It is to be expected that the results will not be better when a GP is confronted in this way with the criteria in a busy practice. Since the justification of decisions is an important issue, it can be concluded that a better registration will result by giving active support. Not only did active support lead to more complete data (as might be expected because a GP is confronted with questions about data which were not entered but were needed for the justification), also the quality of the diagnostic decisions improved. It was shown that a wrong diagnosis was corrected six times after the intervention of FI. Since a GP was confronted with the query belonging to an entered diagnosis, he/she could infer from the questions asked whether a wrong diagnosis was being pursued. After FI's intervention, GPs also were more confident about their diagnoses. Although the number of cases was not large enough to show significant differences in the quality of the diagnoses between all three groups studied, it could be shown that active decision support given to GPs leads to better diagnostic conclusions when compared with the results of GPs who are not or passively supported.

The study made clear that the GPs participating in the experiment were not too familiar with the diagnostic criteria. Only in 66% of the cases, they correctly indicated that the diagnoses fulfilled the criteria.

The cases presented to the GPs were not considered very difficult by them. Therefore, it is the more surprising that the unsupported GPs misdiagnosed still in 11% of the cases and the passively supported GPs made an incorrect diagnosis in 8% of the cases. Perhaps the passively supported GPs did not read the criteria for the diagnosis that well because of the

fact that they considered the cases as relatively easy. However, in daily practice this situation will be encountered many times. The phenomenon that GPs do not take enough notice of criteria which are presented to them was also reported in a study concerning the correctness of problems registered in a regional information system. When incorrectly coded problems were sent back for correction - together with some correct cases -, half of the errors in the problems were not detected by the participating GPs<sup>8</sup>.

Even with active support, still three wrong diagnoses were registered and in four other cases the diagnoses were not justified by the data. In the case of the three wrong diagnoses, it just so happened that for these diagnoses no query was formulated. Therefore, no questions were asked to the GPs. When the criteria would have been translated into queries, a different situation would have existed. In the four other cases, the GPs judged some data not relevant enough.

From the study, it could be concluded that active decision support raises the quality of diagnostic decisions and also results in a more complete data capture.

Nowadays, quality assessment of medical care becomes more and more an important issue. When GPs have to give evidence of whether they provide patient management of quality, the quality of their diagnoses is important. This quality can only be assessed by independent parties if enough corroborating evidence is stored. The study showed that, when active decision support is used, this requirement is fulfilled.

The study also showed that, when criteria are available in free text, these criteria can not be straightforwardly translated into computer-processable queries. The free-text criteria are frequently not explicit enough. Less than half of the available diagnostic criteria and even less of the criteria describing symptoms could be formalized relatively easy.

Although it will be clear to a reader of the criteria that test results should be rather recent (depending on the test), for the software this has to be spelled out in detail. Vague descriptions which may appeal to GPs also have to be translated into explicit descriptions. This can be a time-consuming task.

It was experienced that the Arden Syntax is flexible enough to be used in expressing queries.

## References

1. Lamberts H, Wood M. ICPC - International Classification of Primary Care. Oxford: Oxford University Press, 1987.
2. WONCA. ICHPPC-2-Defined (International Classification of Health Problems in Primary Care). Oxford: Oxford University Press, 1983.
3. Dupuits FMHM, Hasman A. Designing HIOS+. In: Timmers T, Blum BI, eds. Proceedings of the SEMI -- IMIA Working Conference on Software Engineering in Medical Informatics. Amsterdam:

- North-Holland Elsevier Science Publishers BV, 1992:181-193.
4. Dupuits FMHM, Hasman A. Supporting medical decision making. In: Richards B, MacOwan H, eds. *Proceedings of the Current Perspectives in Healthcare Computing Conference 1992 (HC 92)*. Weybridge: BJHC Books, 1992:530-532.
  5. Dupuits FMHM, Hasman A. HIOS+, a decision aid in medicine. In: Lun KC, Degoulet P, Piemme TE, Rienhoff O, eds. *Proceedings of the Seventh World Congress on Medical Informatics 1992 (MEDINFO 92)*. Amsterdam: North-Holland Elsevier Science Publishers BV, 1992:454-460.
  6. Hripcsak G, Clayton PD, Pryor TA, Haug P, Wigertz OB, van der Lei J. The Arden Syntax for Medical Logic Modules. In: Miller RA, ed. *Proceedings of the 14th SCAMC 1990*. Los Alamitos: IEEE Computer Society Press, 1990:200-204.
  7. Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991.
  8. Metsemakers JFM. Unlocking patients' records in general practice for research, medical education and quality assurance: the Registration Network Family Practices, Thesis. Amsterdam: Thesis Publishers, 1994.

### Acknowledgement

Special thanks are expressed to all participating GPs for their involvement, the GP panels for validating the queries and the diagnoses, the Department of Methodology and Statistics of the university for its support in the statistical analyses of the registered data, and especially to the Dutch Ministry of Public Health, Welfare, and Sports (Dutch acronym: VWS; formerly: WVC) for funding the 'ICPC Thesaurus Project'.

## **Chapter 6**

### **User satisfaction of general practitioners with HIOS+, a medical decision support system**

The article presented in this chapter was published in

Dupuits FMHM, Hasman A. User satisfaction of general practitioners with HIOS+, a medical decision support system. *Computer Methods and Programs in Biomedicine* 1995;47:183-188.

The references <sup>3,18,23,24,26,35</sup> included in the List of Publications also concern the research described here.

## **ABSTRACT**

In this article, a user satisfaction study is described. The Formal Interpreter (FI) and the Free-text Formalizer (FF) are modules of the medical DSS HIOS+. With FF, medical experts can formalize topics selected from free-text ICHPPC-II-Defined criteria into queries. FI gives decision support to GPs in their daily work by interpreting queries belonging to entered ICPC codes and presenting advice/warnings concerning the validity (correctness) of the diagnoses in patient cases. In the study, FI was used in field and laboratory test situations by GPs. User satisfaction was assessed using questionnaires consisting of an adapted version of the Bailey questionnaire. Answers given were analysed using Bailey's approach and the 'two sample t test'. The most important result of the FI evaluation was that users given active decision support were more satisfied with FI than users given no or only passive decision support, although only the difference in satisfaction between users given active support and those given passive support was statistically significant.

## **KEYWORDS**

Decision Support System, General Practitioners, Computer User Satisfaction, Queries, ICPC, System Evaluation

## **Introduction**

This article deals with the determination of user satisfaction with a Decision Support System (DSS). A GP information system was evaluated separately and in combination with a DSS. This DSS was developed as part of the national 'ICPC Thesaurus Project' which was initiated to promote a better registration of medical patient data by General Practitioners (GPs). In The Netherlands, medical data are registered in a problem-oriented way using the SOAP approach<sup>1</sup> and the Dutch version of the International Classification of Primary Care (ICPC)<sup>2</sup> for coding complaints, diagnoses, and therapies. The task of the Department of Medical Informatics of the Maastricht University in the project concerned the formalization of free-text criteria belonging to ICPC codes for diagnoses and complaints which are described in the International Classification of Health Problems in Primary Care (ICHPPC)-II-Defined<sup>3</sup>. The task involved an investigation of the usefulness of Decision Support (DS) in the validation of registered ICPC codes on the basis of ICHPPC-II-Defined criteria. The Free-text Formalizer (FF) was developed to enable medical experts to create computer-processable queries by formalizing the free text of ICHPPC-II-Defined criteria. In the experiment, the Formal Interpreter (FI) provides both passive DS by presenting an available ICHPPC-II-Defined criterion belonging to an entered ICPC code, and active DS (advice and warnings) by automatically checking the validity of a registered ICPC code using data in the medical record. In both cases, the DSS intervenes actively.

It is important to determine and quantify user views and expectations concerning a newly developed information system. FI can be regarded typical of a DSS which makes an

investigation of computer user satisfaction the more interesting. If GPs are satisfied with a DSS, they may not only be prepared to adapt their working behaviour in order to take more or full advantage of the Information Technology (IT) provided but may also be prepared to accept DSSs as important tools in daily work. For a successful introduction of DSSs in general practices, it might, therefore, be important to know what GPs think of DS and the IT provided. These aspects were investigated. A first question which had to be answered was whether it is possible to employ an existing user satisfaction method in a situation where GPs are confronted with a specific information system for the first time (although they have a general knowledge about proprietary GP information systems). A second question was whether GPs given active DS by FI will be more satisfied with the system than those given passive or no DS.

In this article, the process of determining user satisfaction of GPs with the FI system is discussed, FI's functioning is described, methods and techniques used in the user satisfaction evaluations of the system are dealt with, results are presented, and a discussion along with a general conclusion is given.

### **The Formal Interpreter**

FI and FF are both modules of the medical DSS HIOS+<sup>4,5,6</sup>, in which the GP registration system HIOS and DS facilities are integrated.

FI gives DS to GPs concerning ICPC code validity. In the active mode, FI activates the corresponding query when a GP enters an ICPC code. FI, then, performs a check concerning the presence of necessary query items in the patient's medical record. If necessary, the GP is questioned about missing items which can be added to the record. When all necessary data are available, FI generates advice regarding the validity of the entered ICPC code<sup>7</sup>. If FI states that the ICPC code is not corroborated, the GP is asked whether the ICPC code needs to be registered.

### **Methods and techniques**

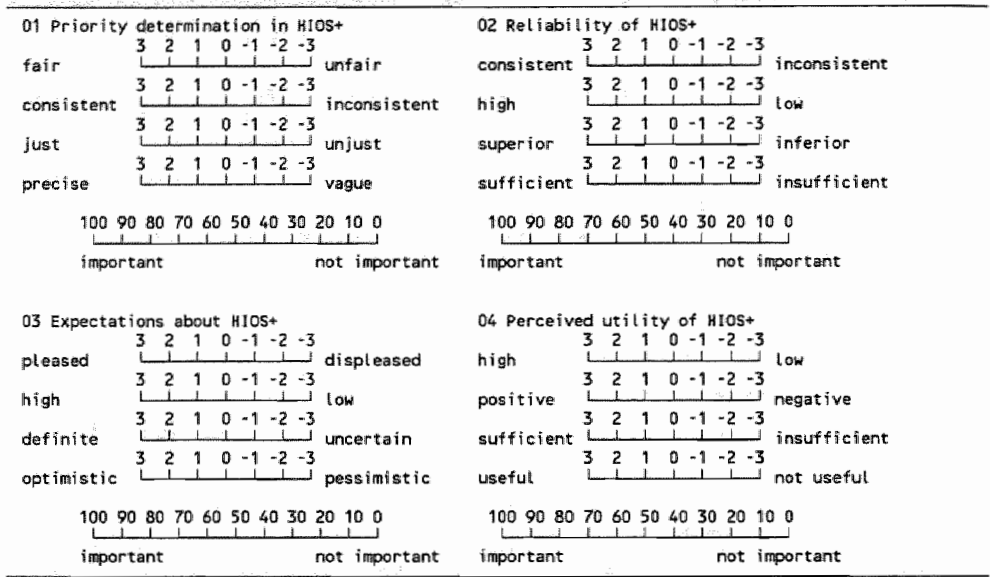
FI was evaluated in a laboratory environment by thirty GPs. In this test, ten true-to-life case descriptions were used. The GPs were asked to enter data concerning the ten cases into HIOS+ after being taught how to work with this system using an eleventh case description.

User satisfaction was assessed using a questionnaire consisting of an adapted version of the Bailey questionnaire<sup>8,9</sup>. Bailey determined a number of factors which are indications of the perceived usefulness of a system. Computer user satisfaction was defined as the sum



of the computer users' weighted (positive/negative) reactions to a set of factors concerning a system employed by the users. For each factor, users express their satisfaction on a scale ranging from -3 to +3 and indicate their opinion about the importance of this factor on a scale ranging from 0% (unimportant) to 100% (important). Figure 1 presents four questions from a questionnaire used in the test situations.

Figure 1. Examples of questions in the Bailey part of the questionnaire



Major aspects covered were: functionality; performance; flexibility; the users' feeling of participation; and integration with the registration system HIOS. Whereas the original Bailey questionnaire consisted of 39 questions, only 27 Bailey questions could be used. Twelve questions had to be excluded because of their irrelevance to the evaluation: they concerned the relationship between users and Electronic Data Processing (EDP) staff and EDP resources/services.

Free-discussion questions added to the FI questionnaire were used to ascertain: (i) if and to what extent the GPs agreed with the ICHPPC-II-Defined criteria used in the case descriptions, (ii) user satisfaction with the case descriptions concerning complexity, clearness, informativeness, realism, and usefulness of automated DS in the described cases, and (iii) the GPs' reactions to the application of active and passive DS in medicine in general.

The users' scores for each satisfaction factor and their measures of importance were used to determine average and normalized (scale ranging from -1 to +1) user satisfaction

scores, taking all factors into account in the way described by Bailey. Using the intervals for normalized user satisfaction scores described by Bailey<sup>9</sup>, user satisfaction could be expressed verbally. Use was made of the 'two sample t test'<sup>10</sup> to determine statistically significant differences between user groups. Answers to the free-discussion questions in the FI questionnaire were interpreted and categorized.

The 30 GPs, who prior to the test were not involved in system design, were divided into three equally large groups. Group I GPs evaluated the usefulness of the GP information system as such. Group II GPs evaluated the system which in addition provided unsolicited textual information. Group III GPs evaluated a full active decision support system.

The question is whether such an evaluation provides the correct satisfaction score when GPs are confronted with a system for the first time. In a field test, therefore, the reactions of three GPs were determined as well.

**Figure 2.** Average user satisfaction scores and t values ( $\alpha = 0.05$ ; t threshold = 2.26; 9 degrees of freedom)

GROUP	USER SATISFACTION SCORES
I	0.64
II	0.42
III	0.78
COMPARISON	t VALUES
I AGAINST II	+ 2.12
I AGAINST III	- 1.10
II AGAINST III	- 3.13

## Results

In this part, FI average and normalized user satisfaction results are described. The average laboratory test scores are presented in the upper part of figure 2. From these average scores, normalized user satisfaction scores of +0.21, +0.14, and +0.26 were computed for the three groups. In accordance with Bailey's intervals, these scores can be described as 'slightly satisfactory'. Group III GPs were more satisfied than group I or group II GPs. Group II GPs were least satisfied. A statistically significant difference existed between groups II and III (figure 2). In figure 3, average user satisfaction scores are shown concerning each one of the 27 original Bailey satisfaction factors.

**Figure 3. Average user satisfaction scores concerning the 27 original Bailey satisfaction factors**

	Group I	Group II	Group III
Priority determination	0.89	0.75	1.09
Expectations	1.04	0.36	0.51
Understanding of the systems	0.90	0.91	0.55
Feeling of control	0.83	0.02	0.83
Relevancy	1.06	0.81	0.72
Perceived utility	0.77	0.65	0.86
Response/turnaround time	-0.72	-0.52	-0.42
Timeliness	0.50	0.40	0.96
Currency	0.42	0.75	0.59
Volume of output	0.49	0.43	0.87
Accuracy	0.73	1.13	1.09
Completeness	0.47	-0.08	0.95
Precision	0.67	-0.02	1.10
Reliability	0.82	0.53	1.09
Confidence in the system	0.62	0.36	0.61
Processing of requests	0.04	0.45	0.81
Effect of the system	0.48	0.03	0.12
Security of data	0.22	0.46	1.37
Flexibility of the system	0.21	-0.10	-0.41
Feeling of participation	0.48	-0.39	-0.08
Documentation of the system	1.27	0.98	1.02
Degree of training	1.19	0.67	1.27
Convenience of access	0.82	0.51	0.84
Format of output	1.18	0.96	0.93
Language	0.81	0.49	1.08
Means of input/output	0.74	0.78	1.18
Integration of the systems	0.74	0.28	1.02

A problem indicated by GPs in all groups concerned the slow response time (question no. 7) which was the result of using an obsolete hardware architecture (a slow network and PCs with 8088 or 80286 CPUs). In groups II and III, low average user satisfaction scores were also obtained for flexibility of the system (question no. 19) and feeling of participation (question no. 20) which resulted from the fact that the GPs were not involved in the development. In addition, the flexibility of the system was restricted intentionally for all groups in order to be able to properly study the impact of DS. In the FI systems used in the study, e.g., it was not possible to deactivate DS, although this option of course does exist in the FI system. Two other satisfaction factors which were valued slightly negatively in group II concerned completeness (question no. 12) and precision (question no. 13) of the

system provided. The result regarding completeness indicated that the provision of only passive support is not satisfactory in a GP environment.

In the field test, FI was used during three months by three GPs in a rural general practice situated near Maastricht. The GPs registered medical patient data during actual encounters, and were advised and warned by FI concerning the validity of registered ICPC codes. User satisfaction with FI was assessed by means of a questionnaire based on Bailey's approach. An average user satisfaction score of +1.02 and a normalized user satisfaction score of +0.34 were obtained. The latter score can be expressed as 'quite satisfactory'.

### **Discussion and general conclusion**

The laboratory test showed that group III GPs were more satisfied with FI than group I or group II GPs. Although there was not a statistically significant difference between groups I and III, it could be concluded that group I GPs were satisfied with the functioning of the GP information system as such and that group III GPs did not experience FI's additional questioning in active decision support as user unfriendly. A statistically significant difference existed between groups II and III. Group II GPs were least satisfied of all. Whereas group I GPs could enter diagnoses in the registration system HIOS without interference from the system and group III GPs were asked by FI about missing items and thus were given supportive feedback, group II GPs were always reminded of the criteria which should be considered in diagnosing the cases. Group II GPs indicated that they often experienced automatic criteria presentation as information overload and therefore hardly read the criteria. This was corroborated by: (i) a time-expenditure analysis which showed that group II GPs averagely spent only about 3 seconds looking at the criteria of all ten cases, and (ii) from results obtained in the free-discussion part of the laboratory test questionnaire. An interview revealed that group II GPs were convinced that they knew enough about the cases which in their opinion were easy and that reading the criteria would not provide new information.

FI laboratory test GPs indicated to be slightly disappointed about DS usefulness in dealing with the cases. The main reason was that routine cases were used, whereas active and passive DS is desired in complex or uncommon cases only (which, by the way, was an interesting result since the GPs were found to misdiagnose even routine cases relatively often<sup>7</sup>). According to the group III GPs, support was not needed in the laboratory test cases. In addition, group II GPs were disappointed about their DS because of the information overload aspect discussed earlier. Despite the GPs' disappointments, the general usefulness of active and passive DS was not questioned. This was concluded from the analysis of the free-discussion questions. In fact, most laboratory test GPs indicated that they would have

liked to receive active and passive DS, and listed a variety of expected advantages.

Regarding the use of Bailey's approach to assess user satisfaction, it should be remarked that Bailey monitored his user group over a relatively long time period whereas the laboratory test was performed in one day and the field test only lasted three months. It is difficult to say in what respects and to what extent the test results would differ from the present ones if FI would have been tested over a longer period of time. Some of the negatively rated factors could be explained easily and were predictable in advance. Slow response times were predictable on the basis of the hardware architecture used. Flexibility and feeling of participation were predictable to be lower due to the GPs' non-involvement in the development of FI as well as the restrictions posed upon the FI systems.

The study showed that significant parts of the Bailey questionnaire could be used. The Bailey factors not used in this evaluation of the information systems are not suitable for a GP environment. The relationship between GPs and the GP information system developer differed from the relation which exists between hospital staff and EDP centres. An interesting remark concerns the many instances in which users gave neutral answers (0 on the scale of -3 to +3). The percentages of zero responses to the 27 original Bailey satisfaction factors came to 19% in group I, 12% in group II, and 7% in group III. The field test of three GPs resulted in a higher satisfaction score. This may indeed indicate that it is not possible to accurately determine the satisfaction score by letting GPs work with a system only for one day. As a result, the laboratory GPs may have been more negative than was necessary.

In conclusion, it can be remarked that not all of the Bailey factors were relevant in the study. Another conclusion is that it will not be possible to rule out mistakes in decision making entirely - not even, if additional DS will be provided -, since human involvement is and will be absolutely necessary in a DSS's functioning. A medical DSS merely functions as a "watchdog" and only is and will be a supportive tool. GPs will always bear the final responsibility for their medical decisions.

Taking into account that the FI evaluations constituted some first exploratory tests, it could be concluded that the way in which the software worked and the degrees of computer user satisfaction were very satisfactory. The test results give good hope that medical DSSs, like HIOS+ FI and FF, in future will become of importance to the medical profession in supporting daily diagnostic and therapeutic decision-making processes.

## References

1. Weed LL. Medical records, medical education and patient care: a problem oriented record as a basic tool. Chicago: Year Book Medical Publishers Inc, 1970.
2. Lamberts H, Wood M. ICPC - International Classification of Primary Care. Oxford: Oxford University

- Press, 1987.
3. WONCA, ICHPPC-2-Defined (International Classification of Health Problems in Primary Care). Oxford: Oxford University Press, 1983.
  4. Dupuits FMHM, Hasman A. Designing HIOS+. In: Timmers T, Blum BI, eds. Proceedings of the SEMI-IMIA Working Conference on Software Engineering in Medical Informatics. Amsterdam: North-Holland Elsevier Science Publishers BV, 1992:181-193.
  5. Dupuits FMHM, Hasman A. Supporting medical decision making. In: Richards B, MacOwan H eds. Proceedings of the Current Perspectives in Healthcare Computing 1992 (HC 92). Weybridge: BJHC Books, 1992:530-532.
  6. Dupuits FMHM, Hasman A. HIOS+, a decision aid in medicine. In: Lun KC, Degoulet P, Piemme TE, Rienhoff O, eds. Proceedings of the Seventh World Congress on Medical Informatics MEDINFO 92. Amsterdam: North-Holland Elsevier Science Publishers BV, 1992:454-460.
  7. Dupuits FMHM, Hasman A. Evaluation of a Decision Support System in a Medical Environment. International Journal of Bio-Medical Computing 1995;38:155-165.
  8. Bailey JE. Development of an Instrument for the Management of Computer User Attitudes in Hospitals. Methods of Information in Medicine 1990;29(1):51-56.
  9. Bailey JE, Pearson SW. Development of a Tool for Measuring and Analyzing Computer User Satisfaction. Management Science 1983;29(5):530-545.
  10. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall, 1991.

### **Acknowledgement**

The researcher involved in the study wishes to thank his colleagues of the Department of Methodology and Statistics for their support in data analysis, all test GPs and (medical) experts for their participation, and especially the Ministry of Public Health, Welfare, and Sports (Dutch acronym: VWS; formerly: WVC) for funding the 'ICPC Thesaurus Project'.



## **Chapter 7**

### **Evaluation of GPs' medical knowledge by a decision support system**

The article presented in this chapter was published in

Dupuits FMHM. Evaluation of GPs' medical knowledge by a decision support system. *Computer Methods and Programs in Biomedicine* 1995;48(1,2):91-96.

The reference <sup>27</sup> in the List of Publications also concerns the research described here.



## **ABSTRACT**

This paper deals with the Evaluator, which is a multipurpose DSS. EV can be applied by creators of automated disease and therapy models to gain insight into knowledge necessary for the diagnosis and the treatment of diseases. GPs, trainees/interns, hospital clinicians, and students can test their medical knowledge with EV. If necessary, EV offers decision support. In a functionality test, EV was used by GPs to assess their knowledge of thyroid disorders. Test conclusions referred to the GPs' level of medical knowledge of thyroid disorders, differences in EV usage by urban and rural GPs, use of thyroid disorder indices, and the GPs' commendations that EV is capable of providing automated decision support to medical professionals in daily practice and that it can very well be used in the education of students and in GP refresher courses.

## **KEYWORDS**

Decision Support System, General Practitioners, System Evaluation, Medical Knowledge

## **Introduction**

The 'Decision Support System in Healthcare' (DSSH) group of the Department of Medical Informatics of the Maastricht University (formerly called the University of Limburg (RL)) focuses its research on automated support in medical decision making of General Practitioners (GPs) especially. A medical Decision Support System (DSS)<sup>1,2</sup> called HIOS+ has been developed to enable the study of various decision support types. HIOS+ offers decision support (using active goal-oriented and data-oriented techniques) to individual users as well as user groups<sup>3,4</sup>. One of its modules is the Evaluator (EV). EV is capable of evaluating medical knowledge and, if necessary, provides decision support. This paper deals with the aims, the development, and the functionality of EV, and describes and discusses results of a functionality test in which EV was used by GPs.

## **Aims and development of the Evaluator**

EV has three aims. First, EV helps creators of automated disease and therapy models to gain insight into knowledge necessary for the diagnosis and the treatment of diseases. By having GPs or hospital clinicians evaluate case descriptions of a specific disease, for instance, a model creator can identify knowledge items which are required or more or less optional for disease management. It can be said that EV helps model creators in gaining insight into the way in which medical professionals reach a diagnosis or determine a therapy of a specific disease. Second, EV enables GPs, trainees/interns, hospital clinicians, and students to test their medical knowledge. EV evaluates case description knowledge items entered by these users on the basis of computerized models of diseases in question. If a user's knowledge proves to be insufficient, EV provides

decision support. Third, EV can be used for education and refresher course purposes.

In EV development, an in-house created Phased Development Methodology (PDM)<sup>5</sup> was used. PDM combines prototyping and other techniques and enables a successful management of development aspects. In the EV knowledge bases, models of all kinds of diseases can be stored. EV is capable of having models evaluated simultaneously by different user groups. By enabling users to import ASCII files with a specified structure, EV guarantees a flexible and extensive application area.

### **Functionality of the Evaluator**

A working process with EV involves: (i) user identification, (ii) model selection, (iii) case description presentation, (iv) knowledge item registration, (v) case description score and missing knowledge item presentation, (vi) repetition of stages iii up to and including v for other case descriptions, (vii) user score presentation concerning the entire model, and (viii) presentation of the 'quit working with EV' option. In the user identification stage, a user has to enter administrative data. In the model selection stage, EV or the user selects one model or a combination of models which is/are to be evaluated. In case model selection is done by the user, he/she is offered a selection menu. After a model has been selected, EV presents case descriptions on the basis of which the model can be evaluated. Each case description represents a specific diagnosis or subdiagnosis. Figure 1 shows a case description for the thyroid disorder struma.

**Figure 1. Example of a case description**

Evaluator V2.02	- Case description -	01-11-1993 02:15:21 pm
CASE DESCRIPTION : Struma (Hyperthyroidism)		
DESCRIPTION :		
Today, a patient who appears to be suffering from the thyroid disorder struma consults you. After having determined the patient's medical history and having performed the physical examination, you refer the patient to the University Hospital Maastricht where one or several laboratory tests will be performed. You have to select and enter necessary data of this patient from and in the screens presented next. In addition, you will have to specify ranges for laboratory test results. In the end, the system will tell you whether your diagnosis of struma (hyperthyroidism) was justified or not.		
F1 = Help      <ENTER> = start evaluation of this case description		

In the knowledge item registration stage, the user first is asked to specify relationships between the case description diagnosis and gender or age. Then, the user selects knowledge items which he/she thinks should be present in the patient if this patient actually has the diagnosed disease. This is shown in figure 2. The user also can select laboratory tests, in which case he/she has to specify ranges for laboratory test results.

Figure 2. Item selection

Evaluator V2.02    - Model: Thyroid disorders -    01-11-1993 02:16:31 pm

CASE DESCRIPTION :    Hyperthyroidism

Select items

EXOPHTALMOS	Y	TOXIC STRUMA	
DIARRHOEA		PALE SKIN	
OBSTIPATION		HOARSE VOICE	Y
NERVOUSNESS	Y	DRY SKIN	
FATIGUE		MOIST SKIN	
THYROID MURMUR		DRY HAIR	
PREFERENCE FOR COLD		MOIST HAIR	
EXCESSIVE PERSPIRATION		PARAESTHESIA	
EYES RED	Y	REDUCED MOBILITY	
BURNING SENSATION IN EYES	Y	INCREASED MOBILITY	
BLINKING OF THE EYELIDS		PULSE RATE > 90/MIN	Y
DOUBLE VISION		PULSE RATE < 80/MIN	
IRREGULAR PALPITATIONS	Y	TSH TEST	
PAIN WHILE SWALLOWING		FT4 TEST	
FINGER TREMOR		INCREASED APPETITE	

Select the items belonging to the case description

1 more

F1 = Help    F7 = End    PgUp = Page up    PgDn = Page down

EV, then, evaluates registered items on the basis of standardized disease indices, and thus determines the likelihood of the patient having the diagnosed disease. When the user is satisfied with the knowledge items registered, he/she selects the 'end of case description' option. EV, then, computes and presents a user case description score and gives advice on the suitability of the diagnosed disease in the patient's case. In case the score is below a preset threshold, the user is given decision support and is presented an overview of knowledge items which were not registered but could have been needed for disease diagnosis or therapy determination. This is shown in figure 3. Subsequently, the user has to specify if he/she wishes to process another case description. On the basis of all case description scores concerning a model, the user can assess in which cases he/she should refresh his/her medical knowledge. At the end, when all case descriptions in a model have been evaluated by the user, EV presents the user mean score for all case descriptions in this model. This score can be compared with the average mean score of the entire user population concerning this model in order to assess how the user scored in comparison to other users and to identify a lack in or a sufficient knowledge of specific

model diagnoses. In case, in the model selection stage, the user selected a combination of models to be evaluated, he/she is automatically presented the first case description of the next model selected earlier. EV regards the combination of selected models as one session. When all case descriptions in the next model have been evaluated by the user, EV presents the user mean score for all case descriptions in this model as well as a session mean score, which is the average of the user mean scores of the models evaluated so far. The process of evaluating models continues until all models in the combination selected earlier have been dealt with. Then, EV presents an overall session mean score.

Figure 3. Example of a decision support screen

Evaluator V2.02		- Model: Thyroid disorders -	01-11-1993 02:17:17 pm
CASE DESCRIPTION : Hyperthyroidism			
SCORE OVERVIEW			
=====			
Numerical score = 35 %		Your final score is 35 %	
Logical score = 0 %			
Missing items :			
TSH TEST	FATIGUE		
FT4 TEST	TOXIC STRUMA		
THYROID MURMUR	REDUCED MOBILITY		
EXCESSIVE PERSPIRATION	PREFERENCE FOR COLD		
FINGER TREMOR	BURNING SENSATION IN EYES		
INCREASED APPETITE	DOUBLE VISION		
HEART PALPITATIONS	BLINKING OF THE EYELIDS		
Continue with next case description? [Y/N] [Y]			
***** The diagnosis was not justified ! *****			

### Evaluator functionality test

Many literary sources describe methods for testing the functionality of DSSs such as EV<sup>6,7</sup>. EV was tested on its functionality by representatives of one of its user groups.

#### Test environment

EV functioning was tested by seventeen GPs who used it to assess their medical knowledge of thyroid disorders. The thyroid disorder model was chosen for EV evaluation purposes because this model is largely standardized and relatively straightforward. In the model selection stage of working with EV, users were not offered a selection menu; instead EV selected the thyroid disorder model automatically. The model was evaluated on the basis of eight case descriptions on primary hyperthyroidism,

primary hypothyroidism, euthyroidism after suspicion of hyperthyroidism, secondary hypothyroidism, Graves' disease, hypophyseal adenoma, De Quervain's disease, and toxic struma (hyperthyroidism) respectively. These thyroid disorders were chosen as subjects of the case descriptions on the basis of a thyroid disorder model which was built by internists. The use of case descriptions in model evaluation was not a new concept, and was applied before in primary care surroundings<sup>8</sup>. In registered knowledge item evaluation, EV made use of the Newcastle Index<sup>9</sup>, Crooks' Index<sup>10</sup>, and Wayne's Index<sup>11,12</sup>, which are existing indices for thyroid disorders. The GPs came from the southern part of the province of Limburg. Six of them work in GP practices and eleven in health centres. Four GPs have less than 1,800 patients each, five GPs have patient populations differing in size between 2,200 and 2,400 patients, and eight GPs have more than 2,400 patients each. Twelve GPs work in rural GP practices/centres; the others work in urban GP practices/centres. Two GPs had over 30 years, three GPs between 20 and 30 years, six GPs between 10 and 20 years, and another six GPs less than 10 years of work experience.

### Test results

The test showed that the GPs had a reasonably well knowledge of the diagnosis and the treatment of thyroid disorders. In figure 4, the GPs are categorized on the basis of their mean scores concerning all eight case descriptions.

**Figure 4. Mean score categorization of all GPs**

Cat.	<90	90-110	110-130	130-150	150-170	170-190	190-210	210-230	≥ 230
GPs	0	1	1	2	2	2	3	3	3

In case a user scored 100 or more points in a case description evaluation, EV concluded that the diagnosis was justifiable. The higher the score, the more justifiable and likely a diagnosis became. The threshold of 100 points was derived from existing and validated thyroid disorder indices. From figure 4, it can be concluded that all GPs, except probably 1 (category 90-110), averagely scored more than 100 points. In some cases, EV indicated a knowledge deficiency concerning a thyroid disorder; knowledge of such a particular thyroid disorder needed to be refreshed by consulting textbooks. Conclusions drawn from the average mean scores were that: - rural GPs in average scored higher than urban GPs (199 points versus 185 points), - GPs with more than 30 years of work experience scored highest (220 points), - GPs with a work experience varying between 15 and 30 years scored lowest (178 points), - GPs with less than 15 years of work experience reached an

average of 196 points, and - GPs with relatively few patients scored better than GPs with relatively many patients.

Averagely, the GPs spent 233 seconds on reading, processing, and score evaluation per case description. Actual processing (registration of knowledge items) of all eight case descriptions in average took a GP 87 seconds. Whereas urban GPs averagely spent 136 seconds on processing, rural GPs in average needed 61 seconds. A GP averagely spent 19 seconds on studying evaluation scores. Whereas urban GPs in average needed 25 seconds for studying scores, rural GPs averagely needed 15 seconds. Studying scores, furthermore, took a more experienced GP less time than it did a less experienced GP.

Knowledge items which were registered by the GPs are also mentioned in the three thyroid disorder indices referred to earlier. Exceptions are the items 'TSH test' and 'FT<sub>4</sub> test'. These tests are not mentioned in the indices because they were not yet being applied at the time the indices were described. All items presented to the GPs in the menus should be registered at least once during the evaluation of the eight case descriptions. Figure 5 shows a list of all items selected more than 30 times and their exact frequencies in descending order.

**Figure 5.      Frequencies of registered knowledge items**

TSH test	113	Constipation	44
Pulse rate	103	Hoarse voice	42
FT <sub>4</sub> test	86	Pain while swallowing	42
Fatigue	83	Dry hair	41
Excessive perspiration	72	Reduced perspiration	41
Dyspnoea on exertion	69	Reduced mobility	41
Heart palpitations	64	Dry skin	40
Nervousness	63	Jerks of the ankle	39
Weight decreased	62	Blinking of the eyelids	38
Finger tremor	61	Puffiness around eyes	37
Preference for cold	60	Painful struma	37
Exophthalmos	52	Burning sensation in eyes	35
Increased appetite	52	Toxic struma	35
Thyroid murmur	45	Double vision	34

#### *Discussion of the test results*

Because of the fact that the number of GPs participating in the test was relatively small, the process of drawing general conclusions was difficult and had to be undertaken with utmost care.

A remarkable general conclusion which could be drawn from the average mean scores was that GPs with a work experience varying between 15 and 30 years scored lower than GPs with less than 15 years of work experience. A possible explanation of this fact was not found.

Another general conclusion which could be drawn from the mean scores and the processing time was that in this test rural GPs seemed to be more effective in diagnosing thyroid disorders than their urban colleagues. This was deduced from the fact that rural GPs scored higher and needed less time for processing. In other words, rural GPs registered more correct knowledge items in less time.

From the overview of the frequencies of the registered knowledge items, model creators could assess which items were absolutely necessary, optional, or not required at all for thyroid disorder management. On the basis of this information, they could adjust their models or even make case description models for particular thyroid disorders. These new models have to be validated in a new evaluation test.

The fact that the registered knowledge items are mentioned in the indices confirms the usability of these instruments as thyroid disorder models. Because the indices have become outdated (archaic laboratory tests are used), however, it will be worthwhile to investigate if they can be adjusted to modern standards. In September 1993, a research project of the Maastricht University and the University Hospital Maastricht was initiated which aims at developing new prediction instruments for thyroid disorders.

## Conclusion

EV is a multipurpose DSS which can be used in different medical settings and by different user categories. With EV, creators of automated disease and therapy models can determine knowledge necessary for the diagnosis and the treatment of diseases. GPs, trainees/interns, hospital clinicians, and students can test their medical knowledge by evaluating case descriptions with EV. EV, therefore, is an important tool in education and refresher courses. This was indicated by seventeen GPs who used EV to test their medical knowledge of thyroid disorders by evaluating eight case descriptions. The test showed that the GPs had a reasonably well knowledge of thyroid disorder management. Case description scores and mean scores averagely exceeded a preset threshold with which the justifiability of thyroid disorder diagnoses was determined. Differences in case description processing could be observed between rural and urban GPs: rural GPs spent less time on reading, processing, and score evaluation of case descriptions than their urban colleagues, and scored higher. Knowledge items registered by the GPs were also mentioned in the

Newcastle, Crooks', and Wayne's indices which were used as indices for the thyroid disorders evaluated in the case descriptions. This proved that these indices can very well be used as a basis for the development of new prediction instruments which will contain standards now in use in modern medicine. The overall opinion of the test GPs was that EV is a commendable tool which is capable of providing automated decision support to medical professionals in daily practice and which can very well be used in the education of students and GP refresher courses.

## References

1. Sprague Jr RH, Watson HJ. Decision Support Systems - Putting Theory into Practice. Englewood Cliffs: Prentice-Hall International Inc, 1986.
2. Turban E. Decision Support and Expert Systems (Managerial Perspectives). New York: Macmillan Publishing Company, 1988.
3. Dupuits FMHM, Hasman A, Ulrichs EMJJ. Diagnostic and therapeutic value of HIOS+. In: Richards B, MacOwan H, eds. Proceedings of the Current Perspective in Health care Conference 1991 (HC 91). Weybridge: BJHC Books, 1991:316-324.
4. Dupuits FMHM, Hasman A, Jansen PCMHF. Classification within HIOS+. In: Adlassnig K-P, Grabner G, Bengtsson S, Hansen R, eds. Proceedings of the Medical Informatics Europe conference 1991 (MIE 91). Berlin/Heidelberg: Springer-Verlag, 1991:749-753.
5. Dupuits FMHM, Hasman A. Designing HIOS+. In: Timmers T, Blum BI, eds. Proceedings of the SEMI -- IMIA Working Conference on Software Engineering in Medical Informatics. Amsterdam: North-Holland Elsevier Science Publishers BV, 1992:181-193.
6. Nwana HS. The evaluation of an intelligent tutoring system. *Intell Tutor Med* 1990;1(3):117-132.
7. Nykänen P, Chowdhury S, Wigertz O. Evaluation of decision support systems in medicine. *Comp Meth Progr Biomed* 1991;34(2/3):229-238.
8. Lawler FH, Hosokawa MC. Evaluation of Standards of Practice for Primary Care Physicians Using 12 Hypothetical Cases. *Jrn Fam Pract* 1987;24(4):377-383.
9. Gurney C, Owen SG, Hall R, Roth M, Harper M, Smart GA. Newcastle Thyrotoxicosis Index. *The Lancet* 1970;19 December:1275-1278.
10. Crooks J, Murray IPC, Wayne EJ. Statistical Methods Applied to the Clinical Diagnosis of Thyrotoxicosis. *Quart Jrn Med* 1959;New series XXVIII (110-April):211-234.
11. Wayne EJ. Clinical and Metabolic Studies in Thyroid Disease. *Brit Med Jrn* 1960, (Jan 2):1-11.
12. Wayne EJ. Clinical and Metabolic Studies in Thyroid Disease. *Brit Med Jrn* 1960, (Jan 9):78-90.

## Acknowledgements

The author wishes to thank all 17 GPs who participated in this study. It was refreshing to see that so many GPs volunteered for the job of testing the software in their own environment and were willing to invest time without receiving any payment at all.





## **Chapter 8**

### **A rational request behaviour - The development of prediction instruments regarding thyroid function tests in primary care**

The article presented in this chapter has been submitted for publication.

The references <sup>7,25,28-32,36</sup> included in the List of Publications also concern the research described here.

## ABSTRACT

**Reason for the study.** GPs request many unnecessary laboratory tests. Feedback about test request adequacy results in a quantitative reduction and a qualitative improvement of requests. The prospective study was performed to stimulate the rational TSH test request behaviour of GPs by means of prediction instruments which give direct feedback about request adequacy on the basis of pretest probabilities of hyperthyroidism and hypothyroidism using patients' signs/symptoms.

**Main findings.** Instruments were created for clinical/subclinical hyperthyroidism and clinical hyperthyroidism using stepwise logistic regression on diagnostic data acquired through questionnaires answered by 80 GPs and 668 patients for whom GPs requested a TSH test.

**Principal conclusion and clinical implications.** At low pretest probability thresholds, the instruments can determine TSH test request adequacy and stimulate a rational TSH test request behaviour of GPs. Use of the instrument for clinical/subclinical hyperthyroidism at a 5% probability threshold in principle can result in a reduction of unnecessary TSH test requests by 37%. With the instrument for clinical hyperthyroidism at a 5% probability threshold, the number of unnecessary test requests in principle can be reduced by 57%.

## KEYWORDS

Hyperthyroidism, Medical Models, Prediction Instruments, Family Medicine

## Introduction

In five out of a hundred patient encounters, a Dutch General Practitioner (GP) requests diagnostic tests which can not be performed in a general practice<sup>1</sup>. Such a request relates to a laboratory test (a chemical blood test: 2 out of 100 patients or a haematologic analysis: 2.1 out of 100 patients), diagnostic imaging (an X-ray: 2.1 out of 100 patients), or another kind of diagnostic test (like a function test: 1.7 out of 100 patients)<sup>1</sup>. For having a test performed, a GP makes use of the facilities of a hospital or a general practice laboratory in the vicinity. Data of the Dutch National Hospital Institute<sup>2</sup> show that the numbers of laboratory and radiologic tests requested by GPs have clearly risen annually. Often it is a question of unnecessary diagnostics<sup>3-5</sup>.

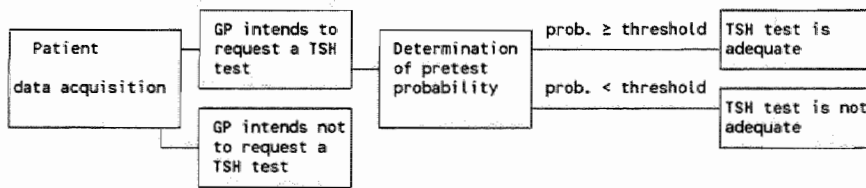
Throughout The Netherlands, several attempts<sup>6-8</sup> have been made to rationalize the diagnostic test request behaviour of GPs. The efforts included the development of a problem-oriented test request form<sup>6</sup>, within which specified tests are advised in case of specific signs, symptoms, and clinical pictures. If only this form is used, there is a risk that too little attention is paid to the previous phase in the diagnostic process. In this phase, it is decided whether there are grounds for further diagnostics on the basis of an estimated pretest probability of a disease, i.e. a pathologic entity which, when confirmed or excluded, has consequences for disease management. Diagnostic and therapeutic

probability thresholds are factors in this decision making.

Twice a year since 1985, the Diagnostic Coordinating Centre (DCC) of the University Hospital Maastricht has been giving individual feedback to about 87 GPs in the Maastricht region regarding their diagnostic test request behaviour in a randomly selected month. This feedback is based on work procedures relating to the use of specific diagnostic tests and on principles of medical decision making. These principles include the estimation of a pretest probability of the presence or absence of a disease, which is an important factor in the decision making. The feedback given by the DCC to the GPs was found to qualitatively improve their request behaviour<sup>9</sup>. Next to this, it proved possible to cut health care expenditure<sup>10-12</sup> by substantially reducing the numbers of requested diagnostic tests<sup>10-12</sup> through giving feedback.

A pretest probability of a disease is estimated on the basis of a patient's data regarding age, gender, medical history, and physical examination. In primary care, data are often missing which are essential in the calculation of pretest probabilities of specified clinical pictures. This, for instance, is the case with thyroid disorders. The past years, the number of Thyroid-Stimulating Hormone (TSH) tests requested by the Maastricht region GPs has increased. An analysis shows that one single sign/symptom or a combination of a few atypical signs/symptoms often lies at the basis of such requests. In most cases, the test outcome is normal.

**Figure 1.** Use of the prediction instruments in TSH test ordering



In order to stimulate a rational TSH test request behaviour of GPs, it was decided to investigate which combinations of signs/symptoms are predictive of the diagnoses of hyperthyroidism or hypothyroidism. A prospective study was initiated, which aimed at supporting the decision-making process of GPs regarding thyroid function test requests by means of prediction instruments with which pretest probabilities of hyperthyroidism or hypothyroidism for patients for whom GPs want to request a TSH test can be determined using data (signs/symptoms) from the primary care setting. In case a prediction instrument indicates that the pretest probability of hyperthyroidism or hypothyroidism is

higher than a certain threshold, a TSH test should be requested. This means that, in this case, the GP's TSH test request is adequate. In figure 1, the decision-making process is presented.

This paper deals with the developed prediction instruments and their results. In addition, the patient groups and methods are described.

## Patients and methods

### Patients

The creation of prediction instruments was based on information about patients for whom a TSH test was requested (either as a single test or as part of a series of tests) by 80 participating Maastricht region GPs via the DCC from 1 September 1993 until 1 July 1994. These patients met the prospective study inclusion criteria of being older than 14 years, not being pregnant, and not being treated or checked up because of thyroid disease. The group comprised 732 patients of whom 64 patients (8.7%) decided not to participate in the study. These 64 patients did not differ from the participating group with respect to age, gender, signs/symptoms, and diagnoses.

Table 1. Description of the patient group

Diagnosis	Reference values		Numbers of patients	
	TSH (mU/l)	FT <sub>4</sub> (pmol/l)	Abs.	%
Clin. hyperth.	≤ 0.2	≥ 24	37	5.5
Subcl. hyperth.	≤ 0.2	10-24	21	3.1
Clin. hypoth.	≥ 5.0	< 10	8	1.2
Subcl. hypoth.	≥ 5.0	10-24	37	5.5
Euthyroidism	0.2-5.0		565	84.6
TOTAL			668	

### Laboratory tests and test outcome interpretation

The 668 patients were classified in various thyroid disorder categories on the basis of their laboratory test outcomes of blood sample analyses. A TSH test and if the TSH ≤ 0.2 or ≥ 5.0 mU/l also a Free Thyroxine (FT<sub>4</sub>) test were performed at the Clinical Chemistry Laboratory of the University Hospital Maastricht using the TSH IRMA and the

Amersham (Map) Analog Method respectively. The various thyroid diagnoses with the corresponding ranges of laboratory test outcomes are shown in table 1. TSH test outcomes ranging from 0.2 to 5.0 mU/l indicate euthyroidism; in these cases, a FT<sub>4</sub> test is superfluous. A TSH test and a FT<sub>4</sub> test cost \$ 25 each.

### *Questionnaires*

In order to acquire signs/symptoms (items) which could be used in the development of the thyroid disorder prediction instruments, GPs and patients were asked to complete questionnaires specific to these target groups.

The GP questionnaire contained a small number of questions about a probable diagnosis and a patient's physical examination, thyroid disorder history, and use of medication. The physical examination part did not comprise questions about items like the thyroid gland size, eyelid lag, and the Achilles' tendon reflex, because these items generally are not examined by GPs, who, therefore, lack routine in reliably determining the (ab)normality of these phenomena.

The patient questionnaire contained 54 questions about signs/symptoms which can attend thyroid disorders. A majority of these signs/symptoms were taken from textbooks and from a number of indices which in earlier years were used clinically to diagnose hyperthyroidism or hypothyroidism<sup>13-18</sup>. The most relevant items are listed in Appendix V (column 1). The patient questionnaire was validated on 221 patients during their visit at a health centre. A GP checked the answers given to the questionnaires. All questions were correctly interpreted by 217 patients (98.2%).

### *Instrument development*

Due to the small number of clinical hypothyroid patients diagnosed during the data acquisition period, reliable prediction instruments could only be created with regard to hyperthyroidism. On the basis of discussions in literature concerning the difference in a clinical sense between clinical and subclinical hyperthyroidism<sup>19-23</sup> diagnosed by means of TSH and FT<sub>4</sub> tests, two prediction instruments were developed: one instrument for clinical hyperthyroidism and one for the combination of clinical and subclinical hyperthyroidism.

Patient characteristics and signs/symptoms obtained from medical histories and physical examinations first were studied univariately. For each item, the relation with laboratory test outcomes was expressed by calculating the items' raw Odds Ratios (ORs). In this, the following formula was used:  $(\text{sensitivity} \times \text{specificity}) / ((100 - \text{sensitivity}) \times$

(100 - specificity)). For every raw OR, a 95% Confidence Interval (95% CI) was calculated. Furthermore, a bivariate analysis was performed with which correlations between items were investigated. Stepwise logistic regression was performed on the entire patient group in order to determine predictive signs/symptoms: an item was added to an equation if  $p < 0.05$  according to the likelihood ratio test; after an item had been added, the set of selected items was reanalysed and an item with  $p > 0.05$ , if any, was removed.

For instrument items, adjusted ORs and 95% CIs were computed with the formulae  $e^{(b-1.96 \times SE)} \leq OR \leq e^{(b+1.96 \times SE)}$  in which  $b$  is the item's weight in logistic regression,  $SE$  is the Standard Error of  $b$ , and  $e^b$  is the OR. If a computed CI did not agree with the likelihood ratio test, it was changed in order to remove the disagreement; this was done for the item preference for heat, for example.

With a resulting logistic regression equation (a prediction instrument), a score can be obtained for each patient for whom a GP intends to request a TSH test. This score can be transformed into a pretest probability using  $p = (100 \times e^{\text{score}}) / (1 + e^{\text{score}})$ . If the pretest probability exceeds a certain diagnostic probability threshold, a TSH test request is adequate.

Receiver Operating Characteristic (ROC) curves were constructed for various probability thresholds of an instrument. For each threshold, the numbers of false positive and false negative laboratory test outcomes, the sensitivity, the specificity, the positive and negative Predictive Values (PVs+ and PVs-), and the positive and negative Likelihood Ratios (LRs+ and LRs-) were determined. CIs for sensitivity and specificity were assessed using tables<sup>24</sup> if  $n < 100$  or the formula  $p \pm 1.96 \times \sqrt{(p \times (1 - p) / n)}$  if  $n \geq 100$ .

Assuming that costs involved are equally high, the optimal point of a ROC curve is theoretically based on the incidence rate and the ratio between sensitivity and specificity. At this point, the sum of false positive and false negative results is minimal. For clinical reasons, preference can be given either to a threshold with a higher sensitivity (relatively few false negative test outcomes) or to a threshold with a higher specificity (relatively few false positive test outcomes).

### *Instrument evaluation*

Instrument quality was determined using an iterative evaluation process called the Grouped Jackknife Method<sup>25</sup>, which is more thorough than a learning/test sets approach: the entire patient group was divided into a subset of about 111 patients for testing (the test set) and the remainder (the learning set) for building a prediction instrument. This was

repeated six times; with each of the six different learning sets, items were selected and their regression weights determined. It should be noted that such a prediction instrument differed from the corresponding prediction instrument discussed under the previous heading in that it was based on only a part of the patient group and therefore was used only for validation purposes. The overall suitability of an instrument eventually was ascertained by combining all test subset results.

**Table 2.** Prediction instruments: a predictive score is obtained by adding an item's logistic regression coefficient (WEIGHT) to the constant if the item is present

ITEM	WEIGHT	SE	OR	95% CI
<u>CLIN./SUBCLIN. HYPER.</u>				
constant	-3.05			
preference for cold	0.95	0.43	2.58	(1.11-6.01)
preference for heat	-0.67	0.37	0.51	(0.25-0.99)
weight decreased	0.71	0.33	2.03	(1.07-3.88)
dry skin	-0.74	0.36	0.48	(0.24-0.97)
pale skin	0.83	0.32	2.30	(1.22-4.29)
dyspnoea on exertion	1.03	0.34	2.80	(1.44-5.45)
smoker	-0.70	0.34	0.49	(0.26-0.97)
pulse rate < 80/min	-0.45	0.37	0.64	(0.31-0.99)
> 90/min	0.82	0.36	2.26	(1.12-4.60)
<u>CLIN. HYPER.</u>				
constant	-4.18			
preference for cold	1.70	0.49	5.48	(2.10-14.30)
weight decreased	1.42	0.40	4.15	(1.89- 9.06)
dry skin	-1.46	0.52	0.23	(0.08- 0.64)
puffiness around eyes	0.98	0.44	2.66	(1.12- 6.31)
dyspnoea on exertion	1.35	0.46	3.85	(1.57- 9.50)
smoker	-1.15	0.44	0.32	(0.13- 0.75)
pulse rate < 80/min	-0.50	0.50	0.61	(0.23- 0.99)
> 90/min	1.18	0.43	3.26	(1.40- 7.56)

## Results

### *Patients*

The group of participating patients ( $n = 668$ ) comprised 112 males (16.8%) with an average age of 49 (15 to 94 years of age) and 556 females (83.2%) with an average age of 45 (15 to 98 years of age). Of the 668 patients who met the inclusion criteria for



participation in the study, 565 patients (84.6%) were found to have a normal thyroid function and 103 patients (15.4%) to have abnormal TSH test outcomes (see table 1). The 37 clinical hyperthyroid patients (5.5%) and the 21 subclinical hyperthyroid patients (3.1%) were of importance to the study.

### *Instrument development*

Associations between patient characteristics and signs/symptoms on the one hand and laboratory test outcomes on the other were expressed by calculating the items' raw ORs. Frequencies and raw ORs of the most relevant items present in hyperthyroid and euthyroid patients are presented in Appendix V. An item has a higher diagnostic value the more its OR differs from 1.00<sup>26</sup>.

All investigated items, of which many are listed in Appendix V, were used in the multivariate analysis (logistic regression) because of the fact that all raw ORs were relatively low. Items which were selected with logistic regression and, therefore, are predictive of clinical hyperthyroidism or clinical/subclinical hyperthyroidism are presented in table 2. The prediction instruments have several items in common, i.e. preference for cold, weight decreased, dry skin, dyspnoea on exertion, being a smoker, and pulse rate < 80/min or > 90/min. Sensitivities, specificities, PVs+, PVs-, LRs+, and LRs- belonging to the various probability thresholds are listed in table 3.

**Table 3. Score/probability thresholds and related values of the prediction instruments for clinical/subclinical and clinical hyperthyroidism**

Score/prob. threshold	Clinical/subclinical hyperthyroidism							
	Sensitivity (%)	Specificity (%)	PV+ (%)	PV- (%)	FPos	FNeg	LR+	LR-
0.85 (70%)	0	100	-	92	0	55	--	1
0.41 (60%)	4	100	67	92	1	53	22.29	0.97
0.00 (50%)	11	99	35	93	11	49	6.08	0.91
-0.40 (40%)	18	98	38	93	16	45	6.97	0.84
-0.84 (30%)	24	98	38	93	21	42	6.90	0.79
-1.38 (20%)	45	93	33	95	50	30	5.57	0.59
-1.73 (15%)	49	91	32	95	57	28	5.28	0.56
-2.19 (10%)	64	78	21	96	134	20	2.91	0.47
-2.93 ( 5%)	84	55	14	97	274	9	1.87	0.30
-3.17 ( 4%)	89	39	12	98	371	6	1.47	0.28
-3.47 ( 3%)	95	31	11	98	426	3	1.36	0.18
-3.85 ( 2%)	96	14	9	98	527	2	1.12	0.26
-4.59 ( 1%)	100	3	8	100	594	0	1.03	0
-5.62 (0.4%)	100	0	8	-	613	0	--	0

Score/prob. threshold	Clinical hyperthyroidism							
	Sensitivity (%)	Specificity (%)	PV+ (%)	PV- (%)	FPos	FNeg	LR+	LR-
1.39 (80%)	0	100	-	94	0	37	--	1
0.85 (70%)	0	99	0	94	2	37	0	1
0.41 (60%)	5	99	50	95	2	35	17.05	0.95
0.00 (50%)	11	99	57	95	3	33	22.74	0.90
-0.40 (40%)	30	99	55	96	9	26	20.84	0.71
-0.84 (30%)	32	99	55	96	10	25	20.46	0.69
-1.38 (20%)	41	96	39	97	23	22	11.12	0.62
-1.73 (15%)	54	92	28	97	52	17	6.56	0.50
-2.19 (10%)	62	89	25	98	69	14	5.68	0.42
-2.93 ( 5%)	81	74	15	99	164	7	3.12	0.26
-3.17 ( 4%)	81	69	13	99	197	7	2.60	0.28
-3.47 ( 3%)	89	59	11	99	258	4	2.18	0.18
-3.85 ( 2%)	92	55	11	99	286	3	2.03	0.15
-4.59 ( 1%)	95	31	7	99	435	2	1.37	0.17
-4.70 (0.91%)	100	24	7	100	477	0	1.32	0
-7.41 (0.06%)	100	0	6	-	631	0	--	0

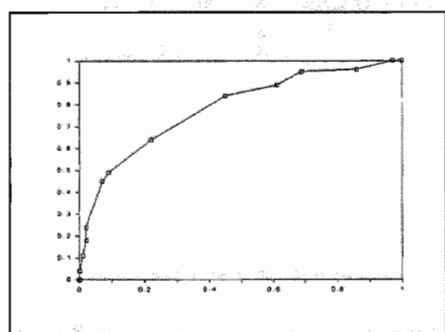
This table additionally contains the numbers of False Positive (FPos) and False Negative (FNeg) predictions determined at each threshold. In this paper's Discussion, attention will be paid to the use of the table 3 values in the choice of thresholds. The ROC curves of the prediction instruments are presented in figure 2.

### Instrument evaluation

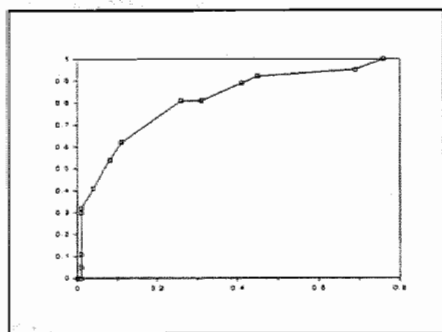
The Grouped Jackknife Method was applied to evaluate both prediction instruments. At a probability threshold of 5%, the instrument for clinical/subclinical hyperthyroidism was found to have a sensitivity of 88% (CI: 77-96%), a specificity of 40% (CI: 36-44%), a PV+ of 12%, a PV- of 97%, a LR+ of 1.46, and a LR- of 0.32. In this case, 7 clinical/subclinical hyperthyroid patients were not detected and a TSH test was performed on 367 euthyroid patients. At a 2% threshold, the instrument was found to have a sensitivity of 95% (CI: 87-99%), a specificity of 14% (CI: 11-17%), a PV+ of 9%, a PV- of 97%, a LR+ of 1.11, and a LR- of 0.38. In this case, 3 clinical/subclinical hyperthyroid patients were not detected and a TSH test was performed on 524 euthyroid patients. Similar results were obtained for the instrument for clinical hyperthyroidism. At a 5% threshold, the instrument's sensitivity, specificity, PV+, PV-, LR+, and LR- came to 89% (CI: 75-97%), 60% (CI: 56-64%), 12%, 99%, 2.22, and 0.18 respectively. In this case, 4 clinical hyperthyroid patients were not detected and a TSH test was performed on 253 euthyroid patients. At a 2% threshold, the instrument's sensitivity, specificity,

PV+, PV-, LR+, and LR- came to 95% (CI: 82-99%), 46% (CI: 42-50%), 9%, 99%, 1.75, and 0.12 respectively. In this case, 2 clinical hyperthyroid patients were not detected and a TSH test was performed on 341 euthyroid patients.

**Figure 2. ROC curves belonging to the instruments for clinical/subclinical and clinical hyperthyroidism (horizontal axis: 1-specificity; vertical axis: sensitivity)**



(a) ROC curve for clinical/subclinical hyperthyroidism



(b) ROC curve for clinical hyperthyroidism

## Discussion

One of the aims of the DCC is to support GPs in their diagnostic test request behaviour with the intention to stimulate an as rational as possible diagnostic request behaviour based on work procedures and principles of medical decision making. The request of diagnostic tests presumes that there is a sufficiently high pretest probability of the presence or absence of a disease based on, for instance, medical history and physical examination data. Diagnostic and therapeutic probability thresholds are factors in this decision making. If the decision is taken to perform further tests, a test is chosen which has a sufficiently high discriminative power with regard to the confirmation or exclusion of a disease.

The estimation of a pretest probability often is not easy due to the fact that the predictiveness of specific signs/symptoms only is known for a limited number of diseases. Examples of diseases with known sign/symptom predictiveness are angina pectoris (coronary artery disease) which is characterized by substernal pain after exertion or emotion and pain relief through rest<sup>27</sup>; infectious mononucleosis which is characterized by

fever, a sore throat, enlarged lymph glands in the throat, splenomegaly, and petechias on the palate<sup>28</sup>; and gallstone disease which is characterized by classic biliary colics<sup>29</sup>. In primary care, not enough research has been carried out into the predictability of signs/symptoms known as being characteristic of hyperthyroidism or hypothyroidism.

When giving feedback to the Maastricht region GPs, the DCC observed that the number of TSH test requests clearly increased over the past years, despite the frequently given feedback about the GPs' TSH test request behaviour. Reasons for the increase are, for instance, that TSH tests are performed as part of a series of tests on patients with vague or general complaints, that patients press their GPs for further laboratory tests, and that thyroid disorders are difficult to diagnose especially in elderly patients. The latter reason relates to the aspect, described in literature<sup>30-35</sup>, that in particular in elderly patients thyroid disorders might manifest themselves monosymptomatically. Relatively often, GPs request a TSH test on the basis of one single sign like fatigue, heart palpitations, weight loss, or excessive perspiration. In most cases, consequently, the test outcome is normal. As a means to stimulate a rational request behaviour with respect to TSH tests, it was decided, in consultation with the participating GPs, to carry out research into the predictiveness of signs/symptoms in the group of patients for whom the GPs requested a TSH test via the DCC.

**Table 4.** Contingency table of the prediction instrument for clinical hyperthyroidism at a probability threshold of 5%

<b>PROB. THRESH. 5%</b>	<b>Hyperthy.</b>	<b>Different outcome</b>	<b>Total</b>	<b>Test characteristic</b>
Number > 5%	33	253	286	PV+ 11.5%
Number ≤ 5%	4	378	382	PV- 99.0%
Total	37	631	668	
Test char.	<b>Sens 89%</b>	<b>Spec 60%</b>		<b>LR+ 2.22 LR- 0.18</b>
Achieved test reduction of 57.2% (patients' probability ≤ 5%)				

The research showed that the ORs of the various signs/symptoms were relatively low. This is understandable because in the research only those patients were analysed who had already been selected by the GPs for a TSH test (on the basis of one or a few

complaints). Test characteristics of the prediction instruments were determined at various probability thresholds of the instruments' ROC curves (see table 3). Use of the prediction instrument for clinical hyperthyroidism at the probability threshold of 5% led (see table 4) to an incidence rate of clinical hyperthyroidism in the selected group (the PV+) of 12% as opposed to an incidence rate of 5.5% in the group on which the creation of the instrument was based. This use at the 5% threshold resulted in a reduction in the number of TSH tests requested with respect to the research group by about 57%. The problem attended with the choice of the 5% probability threshold was the number of undetected clinical hyperthyroid patients: 4 out of 37 patients (11%). At a threshold with a higher sensitivity, the number of eventually missed diagnoses was lower: at the 2% threshold, 2 out of 37 patients (5%) were missed (see table 5).

**Table 5. Contingency table of the prediction instrument for clinical hyperthyroidism at a probability threshold of 2%**

<b>PROB. THRESH. 2%</b>	<b>Hyperthy.</b>	<b>Different outcome</b>	<b>Total</b>	<b>Test characteristic</b>
Number > 2%	35	341	376	PV+ 9.3%
Number ≤ 2%	2	290	292	PV- 99.3%
Total	37	631	668	
Test char.	<b>Sens</b> <b>95%</b>	<b>Spec</b> <b>46%</b>		<b>LR+ 1.75</b> <b>LR- 0.12</b>

Achieved test reduction of 43.7% (patients' probability ≤ 2%)

However, use of the 2% threshold also resulted in a higher number of TSH test requests (a test reduction of 44% instead of 57%). Similar results were obtained for the instrument for clinical/subclinical hyperthyroidism.

In consultation with the GPs, a probability threshold needs to be chosen for each prediction instrument. The seriousness of missing hyperthyroid patients and cost considerations are factors in the discussion concerning threshold choices. The FT<sub>4</sub> concentration expresses the severeness of a thyroid disorder in a patient. Assuming that a FT<sub>4</sub> concentration of more than 30 pmol/l indicates severe clinical hyperthyroidism, 1 patient of the 3 clinical hyperthyroid patients who were not detected with the prediction instrument for this thyroid disorder at a 2% threshold (see table 3) had severe clinical

hyperthyroidism. Of the 7 patients missed at a 5% threshold, 3 had severe hyperthyroidism. The circumstance that severely diseased patients were missed can have implications for the patients' care and community costs involved in the patients' probable future diagnosis and treatment. The nondetection of a clinical hyperthyroid patient, namely, can lead to: (i) a more serious disease, like cardiac arrhythmia, cardiac decompensation, or arrhythmia during operation or narcosis, in the short run, and (ii) increased osteoporosis or physical damage to the heart in the long run. In the framework of the feedback, the chosen thresholds and related matters will be described in work procedures. In individual cases, GPs of course will be permitted to deviate from procedures on reasoned grounds.

The feedback given to GPs regarding their test request behaviour also cuts test-ordering expenditure. In 1985, the year in which feedback was given for the first time, the net saving came to \$ 55,000<sup>12</sup>. This amount was about as high as the cost of providing feedback. The feedback given to GPs about their general test ordering in the past years has resulted in a cost saving of  $\pm$  \$ 369,000 annually. The research described in this paper shows that the application of the prediction instruments may cut expenditure even further. The prevention of unnecessary TSH tests only by applying the instrument for clinical hyperthyroidism, e.g., would have accounted for an additional annual saving of \$ 11,460 ( $382 \times 1.2 \times \$ 25$ ) at a 5% threshold and of \$ 8,760 ( $292 \times 1.2 \times \$ 25$ ) at a 2% threshold. The multiplication factors of 1.2 relate to the conversion of the 10-month results into annual ones.

It can be concluded that on the basis of these research findings it seems useful that GPs apply the prediction instruments in order to determine TSH test request adequacy and, thus, to come to a more rational thyroid disorder test request behaviour. By extending the data acquisition period, more information was collected about clinical hyperthyroid and clinical hypothyroid patients on the basis of which ultimately also a reliable prediction instrument for clinical hypothyroidism could be developed. The results of this research are presented in the Addendum to chapter 8. Similar research as was carried out into the development of prediction instruments for thyroid disorders also can be performed for other diseases frequently occurring in primary care.

## ADDENDUM TO CHAPTER 8

### Introduction

As is apparent from chapter 8, not enough clinical hypothyroid patients were detected in the initial data acquisition period to enable the creation of a prediction instrument for the respective thyroid disorder. To this end, additional information was collected about new clinical hyperthyroid and clinical hypothyroid patients by continuing data acquisition after 1 July 1994. The additional data did not result in an item modification of the prediction instrument for clinical hyperthyroidism presented in chapter 8. On the basis of the additional information about new clinical hypothyroid patients, however, a reliable prediction instrument for clinical hypothyroidism could be developed. The results are presented in this Addendum.

### Patients and methods

From 1 July 1994 until 1 November 1996, additionally 111 new patients were sent by the 80 GPs to the DCC for a TSH test and a subsequent  $FT_4$  test and asked to complete a questionnaire. Ninety of these patients had clinical hyperthyroidism or clinical hypothyroidism. Data about the remaining 21 patients were obtained due to a change in applied laboratory ranges during the acquisition period and transitional difficulties resulting from acquiring data on all possible thyroid patients to selective data acquisition on only two thyroid disease categories. The patients were subjected to the same study inclusion criteria and the same methods of data acquisition and analysis as were the patients discussed in chapter 8.

The selective acquisition of data about only two patient categories had consequences for the use of logistic regression in the development process. Due to the artificially changed incidence rates in the patient group, the instrument's constant  $\beta_0$  - which is automatically determined by logistic regression - had to be corrected through recalculation using the formula<sup>36</sup>:  $\beta_{0new} = \beta_{0old} + \ln(\text{inc}_{new}/(1-\text{inc}_{new})) - \ln(\text{inc}_{old}/(1-\text{inc}_{old}))$  in which inc is the incidence rate, old refers to the 10 months' patient group, and new relates to the cumulative group of 779 patients.

### Results

#### Patients

The cumulative patient group ( $n = 779$ ) comprised 126 males (16.2%) with an average age of 52 (ranging from 15 to 96 years of age) and 653 females (83.8%) with an

average age of 48 (ranging from 15 to 98 years of age). Of these 779 patients, 576 patients (73.9%) were found to have a normal thyroid function and 203 patients (26.1%) to have abnormal TSH test outcomes. The disease categorization of the cumulative patient group is shown in table 6.

**Table 6.** Description of the cumulative patient group

Diagnosis	Cumulative numbers of patients	
	Abs.	%
Clinical hyperthyroidism	87	11.2
Subclinical hyperthyroidism	22	2.8
Clinical hypothyroidism	48	6.2
Subclinical hypothyroidism	46	5.9
Euthyroidism	576	73.9
Total	779	

The data of the 87 clinical hyperthyroid patients (11.2%) and the 48 clinical hypothyroid patients (6.2%) were investigated.

#### *Instrument development*

Frequencies and raw ORs of the most relevant items present in clinical hypothyroid and euthyroid patients are presented in Appendix VI.

A result of the univariate analysis was that one item was identified of whom all hypothyroid patients suffered. This concerned the item of fatigue. It was found that patients who did not experience fatigue did not have hypothyroidism. On the basis of this result, fatigue is applied as a prerequisite for the use of the prediction instrument for clinical hypothyroidism. Items which were selected with logistic regression and therefore are predictive of clinical hypothyroidism are presented in table 7.

Because of its medical relevance, the item cluster of preference for heat, dry skin, and dry hair was offered in the multivariate analysis and subsequently was selected as a relevant item in logistic regression. The item of swollen throat was found twice, in that painless swollen throat and painful swollen throat were both selected with logistic regression as being positively indicative of clinical hypothyroidism with the difference that the first-mentioned item contributes more to the probability of the diagnosis than the latter.



**Table 7. The prediction instrument for clinical hypothyroidism**

ITEM	WEIGHT	SE	OR	95% CI
<b>CLIN. HYPO.</b>				
constant	-3.50			
irritable	-1.05	0.34	0.35	(0.18- 0.68)
reduced mobility	0.79	0.35	2.21	(1.11- 4.38)
weight increased	0.83	0.33	2.29	(1.20- 4.38)
finger tremor	-1.03	0.37	0.36	(0.17- 0.74)
puffiness around eyes	0.71	0.35	2.02	(1.02- 4.04)
painful swollen throat	0.15	0.59	1.16	(1.01- 3.69)
painless swollen throat	1.08	0.38	2.95	(1.40- 6.20)
(preference for heat, dry skin, dry hair)	0.96	0.42	2.62	(1.15- 5.95)

Sensitivities, specificities, PVs+, PVs-, LRs+, LRs-, and the numbers of False Positive (FPos) and False Negative (FNeg) predictions belonging to the various probability thresholds of the instrument are listed in table 8.

**Table 8. Score/probability thresholds and related values of the prediction instrument for clinical hypothyroidism**

Score/prob. threshold	Sensitivity (%)	Clinical hypothyroidism						LR+	LR-
		Specificity (%)	PV+ (%)	PV- (%)	FPos	FNeg			
0.41 (60%)	0	100	-	94	0	48	--	1	
0.00 (50%)	4	100	100	94	0	46	--	0.96	
-0.40 (40%)	17	99	62	95	5	40	24.37	0.84	
-0.84 (30%)	21	99	53	95	9	38	16.92	0.80	
-1.38 (20%)	35	94	29	96	42	31	6.16	0.69	
-2.19 (10%)	60	83	19	97	122	19	3.62	0.48	
-2.93 ( 5%)	83	65	13	98	259	8	2.35	0.26	
-3.85 ( 2%)	98	40	10	99	441	1	1.62	0.05	
-4.59 ( 1%)	98	31	9	99	502	1	1.43	0.07	
-5.27(0.51%)	100	13	7	100	637	0	1.15	0	

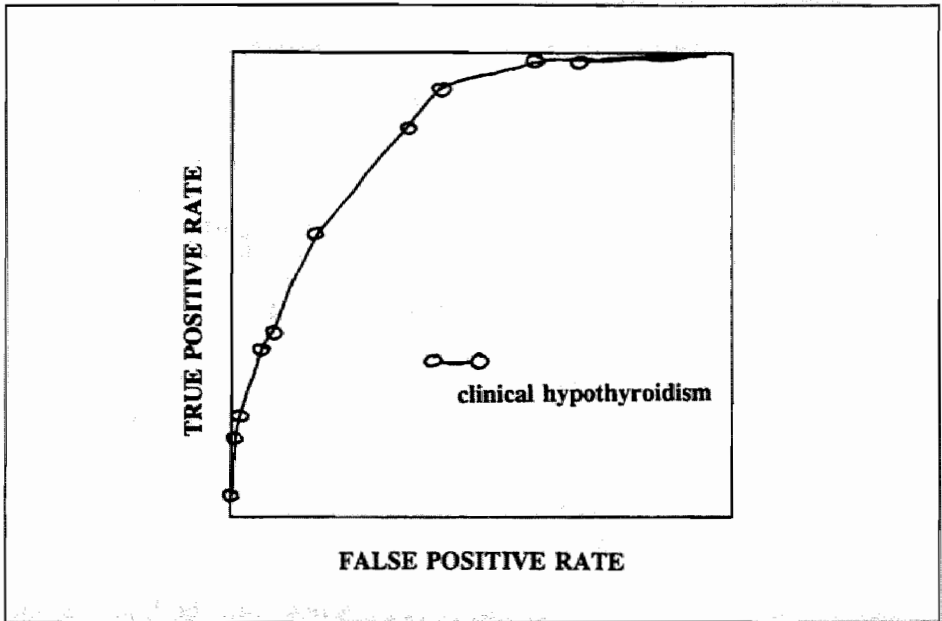
The ROC curve of the prediction instrument is presented in figure 3.

### *Instrument evaluation*

Instrument quality was determined with the Grouped Jackknife Method in which seven times about 111 different patients from the cumulative patient group of 779 new patients were used for testing purposes and the remainder for instrument creation. At a

probability threshold of 5%, the instrument for clinical hypothyroidism was found to have a sensitivity of 79% (CI: 65-90%), a specificity of 62% (CI: 58-66%), a PV+ of 12%, a PV- of 98%, a LR+ of 2.07, and a LR- of 0.34. In this case, 10 clinical hypothyroid patients (21%) were not detected and a TSH test was performed on 280 euthyroid patients (38%). At a 2% threshold, the instrument was found to have a sensitivity of 90% (CI: 77-97%), a specificity of 46% (CI: 42-50%), a PV+ of 10%, a PV- of 99%, a LR+ of 1.66, and a LR- of 0.23. In this case, 5 clinical hypothyroid patients (10%) were not detected and a TSH test was performed on 395 euthyroid patients (54%).

**Figure 3.** The ROC curve of the prediction instrument for clinical hypothyroidism



### Discussion

For use in daily practice, one probability threshold needs to be chosen for the prediction instrument for clinical hypothyroidism. The choice depends on a GP's point of view concerning the seriousness of missing clinical hypothyroid patients and cost considerations. At a 2% threshold, 1 patient was missed with the prediction instrument

(see table 8). This patient had a FT<sub>4</sub> concentration of about 6.2 pmol/l which can be described as nonsevere clinical hypothyroidism (the assumed threshold for severe illness is 4 pmol/l). Of the 8 patients missed at a 5% threshold, none had severe hypothyroidism. Use of the instrument at the probability threshold of 5% led to an incidence rate of clinical hypothyroidism in the selected group (the PV+) of 11.9% as opposed to an incidence rate of 6.2% in the group on which the creation of the instrument was based.

**Table 9. Contingency table of the prediction instrument for clinical hypothyroidism at a probability threshold of 5%**

PROB. THRESH. 5%	Hypothy.	Different outcome	Total	Test characteristic
Number > 5%	38	280	318	PV+ 11.9%
Number ≤ 5%	10	451	461	PV- 97.8%
Total	48	731	779	
Test char.	Sens 79%	Spec 62%		LR+ 2.07 LR- 0.34

Achieved test reduction of 59.2% (patients' probability ≤ 5%)

This is shown in table 9. Use of the instrument at the 5% threshold resulted in a reduction in the number of TSH tests requested with respect to the research group by about 59%. The problem attended with the choice of the 5% probability threshold was the number of undetected clinical hypothyroid patients: 10 out of 48 patients (21%). At a threshold with a higher sensitivity, the number of eventually missed diagnoses was lower: at the 2% threshold, 5 out of 48 patients (10%) were missed. This is shown in table 10. However, use of the instrument at the 2% threshold also resulted in a higher number of TSH test requests (a test reduction of about 44% instead of 59%). Generally, the observation could be made that more clinical hypothyroid than clinical hyperthyroid patients were missed if a higher threshold was chosen in the Grouped Jackknife evaluation.

The prediction instrument for clinical hypothyroidism may also account for an additional cut in expenditure. The prevention of unnecessary TSH tests only by applying the instrument would have accounted for an additional annual saving of \$ 3,688 (461 x 0.32 x \$ 25) at a 5% threshold and of \$ 2,728 (341 x 0.32 x \$ 25) at a 2% threshold.

**Table 10. Contingency table of the prediction instrument for clinical hypothyroidism at a probability threshold of 2%**

<b>PROB. THRESH. 2%</b>	Hypothy.	Different outcome	Total	Test characteristic
Number > 2%	43	395	438	PV+ 9.8%
Number ≤ 2%	5	336	341	PV- 98.5%
Total	48	731	779	
Test char.	<b>Sens 90%</b>	<b>Spec 46%</b>		<b>LR+ 1.66 LR- 0.23</b>

Achieved test reduction of 43.8% (patients' probability ≤ 2%)

The created, separate prediction instruments for clinical hyperthyroidism, clinical/subclinical hyperthyroidism, and clinical hypothyroidism were developed on the basis of the patients who had TSH and FT<sub>4</sub> test outcomes indicative of these thyroid diagnoses. The instruments can without difficulty be incorporated in Medical DSSs (MDSSs), like HIOS+<sup>37,38</sup> developed at the Maastricht University. A MDSS can process the instruments automatically or one of them on the basis of a GP's instruction. In addition to the separate instruments, also a combination of both clinical instruments was determined. The test characteristics of this combination are presented in table 11.

**Table 11. Score/probability thresholds and related values of the combination of the prediction instruments for clinical hyperthyroidism and clinical hypothyroidism**

Score/prob. threshold	Clinical combination model							
	Sensitivity (%)	Specificity (%)	PV+ (%)	PV- (%)	FPos	FNeg	LR+	LR-
1.39 (80%)	0	100	-	83	0	135	--	1
0.85 (70%)	0	99	0	83	2	135	0	1
0.41 (60%)	3	99	67	83	2	131	9.54	0.97
0.00 (50%)	10	99	82	84	3	121	22.26	0.90
-0.40 (40%)	21	98	67	86	14	106	9.88	0.80
-0.84 (30%)	26	97	65	86	19	100	8.79	0.76
-1.38 (20%)	39	90	46	88	62	82	4.08	0.67
-1.73 (15%)	50	84	40	89	102	68	3.13	0.60
-2.19 (10%)	61	73	32	90	175	52	2.26	0.53
-2.93 (5%)	85	45	24	94	356	20	1.54	0.33

(continued)

Score/prob. threshold	Sensitivity (%)	Specificity (%)	PV+ (%)	PV- (%)	FPos	FNeg	LR+	LR-
-3.17 ( 4%)	92	39	24	96	396	11	1.49	0.21
-3.47 ( 3%)	93	33	28	96	430	10	1.39	0.22
-3.85 ( 2%)	95	22	20	95	500	7	1.22	0.23
-4.59 ( 1%)	96	15	19	95	546	5	1.14	0.24
-5.27(0.51%)	100	12	19	100	569	0	1.13	0
-7.41(0.06%)	100	0	17	-	644	0	1	--

A comparison of table 12 with tables 5 and 10 shows that the combination results in a higher PV+, a lower PV-, a lower specificity, and a lower test reduction.

**Table 12.** Contingency table of the combination of the prediction instruments for clinical hyperthyroidism and clinical hypothyroidism at a probability threshold of 2%

PROB. THRESH. 2%	Thyroid disorders	Different outcome	Total	Test characteristic
Number > 2%	128	500	628	PV+ 20.4%
Number ≤ 2%	7	144	151	PV- 95.4%
Total	135	644	779	
Test char.	Sens 95%	Spec 22%		LR+ 1.22 LR- 0.23

Achieved test reduction of 19.4% (patients' probability ≤ 2%)

In theory, it is also possible to develop prediction instruments based on the GPs' diagnostic hypotheses. However, the clinical pictures of thyroid patients are not clearly recognizable in general practice situations due to vague signs/symptoms which are uncharacteristic of hyperthyroidism or hypothyroidism. In practice, therefore, a GP mostly has no specific diagnosis in mind and requests the TSH test as part of a series of tests. A differentiation of the hypotheses regarding clinical hyperthyroidism and clinical hypothyroidism in instrument development, however, is expected to lead to higher PVs+ as test results.

It can be concluded on the basis of the research findings concerning the prediction instrument for clinical hypothyroidism that this instrument is useful to GPs in determining

TSH test request adequacy and thus in coming to a more rational thyroid disorder test request behaviour.

## References

1. Groenewegen PP, de Bakker DH, van der Velden J. Nationale Studie naar Ziekten en Verrichtingen in de Huisartspraktijk - Basisrapport: Verrichtingen in de Huisartspraktijk. Utrecht: Nederlands Instituut voor Onderzoek van de Eerstelijnsgezondheidszorg (NIVEL), 1992.
2. Nationaal Ziekenhuis Instituut. De Intramurale Gezondheidszorg in Cijfers. Utrecht: Nationaal Ziekenhuis Instituut (NZI), 1991.
3. Bernstein RM, Hollingworth GR, Wood WE. Prompting physicians for cost-effective test ordering in the low prevalence conditions of family medicine. *Proc Ann Symp Comput Appl Med Care*. 1994;824-828.
4. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for the outpatient diagnostic tests. *N Engl J Med*. 1990;322:1499-1504.
5. Dowling PT, Alfonsi G, Brown MI, Culpepper L. An education program to reduce unnecessary laboratory tests by residents. *Acad Med*. 1989;64:410-412.
6. van Geldrop WJ, Lucassen PLBJ, Smithuis LOMJ. Een probleemgeoriënteerd aanvraagformulier voor laboratoriumonderzoek - Effecten op het aanvraaggedrag van huisartsen. *Huisarts en Wetenschap*. 1992;35:192-196.
7. Pop P, Winkens RAG. Periodieke persoonlijke feedback over het diagnostisch handelen - Ervaren uit het Diagnostisch Coördinerend Centrum Maastricht. *Huisarts en Wetenschap*. 1989;32:205-207.
8. den Ottolander GJH. Het effect van persoonlijke feedback op het aanvraagpatroon van huisartsen. *Medisch Contact*. 1994;75-79.
9. Winkens RAG, Pop P, Bugter-Maessen AMA, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *The Lancet*. 1995;345:498-502.
10. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J*. 1992;304:1093-1096.
11. Winkens RAG, Pop P, Bugter-Maessen AMA, et al. Effects of routine individual feedback over nine years on general practitioners' requests for tests. *Br Med J*. 1996;312:490.
12. Winkens RAG, Ament AJHA, Pop P, Reniers PHA, Grol RPTM, Knottnerus JA. Routine individual feedback on requests for diagnostic tests: An economic evaluation. *Med Dec Mak*. 1996;16:309-314.
13. Crooks J, Murray IPC, Wayne EJ. Statistical methods applied to the clinical diagnosis of thyrotoxicosis. *Quart J Med*. 1959;XXVIII(110):211-234.
14. Crooks J, Wayne EJ, Robb RA. A clinical method of assessing the results of therapy in thyrotoxicosis. *Lancet*. 1960:397-401.
15. Gurney C, Owen SG, Hall R, Roth M, Harper M, Smart GA. Newcastle thyrotoxicosis index. *Lancet*. 1970:1275-1278.
16. Wayne EJ. Clinical and metabolic studies in thyroid disease (The first of two Lumleian Lectures delivered before the Royal College of Physicians of London on April 14 and 16, 1959). *Br Med J*. 1960:1-11.
17. Billewicz WZ, Chapman RS, Crooks J, et al. Statistical methods applied to the diagnosis of hypothyroidism. *Quart J Med*. 1969;XXXVIII(150):255-266.
18. Wayne EJ. Clinical and metabolic studies in thyroid disease (The second of two Lumleian Lectures delivered before the Royal College of Physicians of London on April 14 and 16, 1959). *Br Med J*. 1960:78-90.
19. Bommer M, Eversmann T, Pickardt R, Leonhardt A, Naber D. Psychopathological and neuropsychological symptoms in patients with subclinical and remitted hyperthyroidism. *Klin Wochenschr*. 1990;68(11):552-558.

20. Mudde AH. Effects of Subclinical Hyperthyroidism on Bone and Heart [Doctoral Thesis at the Maastricht University]. Maastricht: University Press Maastricht (UPM), 1993.
21. Ross DS, Neer RM, Ridgway EC, Daniels GH. Subclinical hyperthyroidism and reduced bone density as a possible result of prolonged suppression of the pituitary-thyroid axis with L-thyroxine. *Am J Med.* 1987;82:1167-1170.
22. Schlote B, Schaaf L, Schmidt R, et al. Mental and physical state in subclinical hyperthyroidism - Investigations in a normal working population. *Biol Psych.* 1992;32(1):48-56.
23. Wiersinga WM. Subclinical hypothyroidism and hyperthyroidism - I. Prevalence and clinical relevance. *Neth J Med.* 1995;46(4):197-204.
24. Lentner C. Geigy Scientific Tables - Volume 2: Introduction to Statistics, Statistical Tables, Mathematical Formulae. Basel: Ciba-Geigy Limited, 1982.
25. Miller RG. The Jackknife - A review. *Biometrika.* 1974;61:1-15.
26. Kraemer HC. Evaluating Medical Tests - Objectives and Quantitative Guidelines. London: Sage, 1992.
27. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. *N Engl J Med.* 1979;300:1350-1358.
28. Aronson MD. Heterophil antibody in adults with sore throat -Frequency and clinical presentation. *Ann Int Med.* 1982;96:505-508.
29. Griner PF. Cholelithiasis. In: Griner PF, Panzer RJ, Greenland P, eds. Clinical Diagnosis and the Laboratory: Logical Strategies for Common Medical Problems. Chicago: Chicago Year Book Medical Publishers, 1986:247-255.
30. Bagchi N, Brown TR, Parish RF. Thyroid dysfunction in adults over age 55 years - A study in an urban US community. *Arch Intern Med.* 1990;150:785-787.
31. Davies AB, Williams I, John R, Hall R, Scanlon MF. Diagnostic value of thyrotropin releasing hormone tests in elderly patients with atrial fibrillation. *Br Med J.* 1985;291:773-776.
32. Finucane P, Rudra T, Church H, et al. Thyroid function tests in elderly patients with and without an acute illness. *Age and Ageing.* 1989;18:398-402.
33. Francis T, Wartofsky L. Common thyroid disorders in the elderly. *Postgrad Med.* 1992;92(3):225-230+233-236.
34. Levy EG. Thyroid disease in the elderly. *Med Clin North Am.* 1991;75(1 - Special on Thyroid Diseases):151-167.
35. Sawin CT, Castelli WP, Hershman JM, McNamara P, Bacharach P. The aging thyroid - Thyroid deficiency in the Framingham study. *Arch Intern Med.* 1985;145:1386-1388.
36. Schouten HJA. Klinische statistiek - Een praktische inleiding in methodologie en analyse. Houten: Bohn, Stafleu & van Loghem, 1995.
37. Dupuits FMHM. The use of the Arden Syntax for MLMs in HIOS+, a decision support system for general practitioners in The Netherlands. *Comp Biol Med.* 1994;24(5):405-410.
38. Dupuits FMHM, Hasman A. Evaluation of a decision support system in a medical environment. *Int J Bio-Med Comp.* 1995;38:155-165.

## Acknowledgements

Thanks are expressed to the 80 Maastricht region GPs, the patients, the DCC desk staff, the consulted internists, and the clinical chemist for participating in the study. It was highly appreciated that so many GPs took part. Patients are thanked for conscientiously completing the long questionnaire. Gratitude is expressed to the DCC desk staff for their enormous input and unremitting effort. Finally, Dr. H.J.A. Schouten is thanked for his support, encouragement, and practical knowledge in the statistical analyses performed throughout the entire research.

## **Chapter 9**

### **Hyperthyroidism and hypothyroidism in outpatient and inpatient clinics**

The article presented in this chapter has been submitted for publication.



## ABSTRACT

**Background.** Fewer thyroid patients were detected by GPs during the 10 months' prediction instrument study than could be expected from incidence rates. This suggests that GPs diagnostically miss a number of thyroid patients.

**Aims.** A retrospective study was initiated to investigate: the nondetection rate of new clinical hyperthyroid and clinical hypothyroid patients by the GPs (i.e. the hospital detection rate of these patients) over the 10 months, to what extent the GPs could have diagnosed these patients, and how and where the diagnoses were established in hospital.

**Methods.** The detection rate and places of detection were assessed through case finding with the laboratory information system. Detection methods were investigated using medical records and discharge letters of thyroid inpatients and outpatients. An internist assessed the GPs' diagnostic (in)adequacy using medical records and GPs' referral letters.

**Results.** Major observations were that more clinical thyroid patients were detected in hospital than in primary care, that the nondetection rate of these patients by the GPs was high, and that most patients were referred for nonthyroid diseases. It could not be assessed conclusively why hospital-diagnosed thyroid patients had not been detected by the GPs. Patients' clinical pictures may have been insufficiently clear.

**Conclusion.** The nondetection of thyroid patients in primary care is worthwhile investigating in a future prospective study, in which the prediction instruments for thyroid disorders may be of use.

## KEYWORDS

Thyroid Disorders, Family Medicine, Symptomatology, Decision Making

## Introduction

The development of prediction instruments regarding the diagnoses of hyperthyroidism and hypothyroidism was partly based on the group of patients who were sent by 80 General Practitioners (GPs) for a Thyroid-Stimulating Hormone (TSH) test via the Diagnostic Coordinating Centre (DCC) to the University Hospital Maastricht during the period of 1 September 1993 until 1 July 1994. Criteria of being older than 14 years, not being pregnant, and not taking thyroid medication were applied to these patients to determine their suitability for inclusion in the study.

During the 10 months, the following thyroid disorders were diagnosed in the group of 668 patients: 37 x clinical hyperthyroidism, 21 x subclinical hyperthyroidism, 8 x clinical hypothyroidism, 37 x subclinical hypothyroidism, and 565 x euthyroidism. Since the entire group to which these patients belong is estimated to comprise about 176,000 persons, the deduced incidence rates come to 25 clinical hyperthyroid, 40 clinical/subclinical hyperthyroid, 6 clinical hypothyroid, and 31 clinical/subclinical hypothyroid patients in a population of 100,000 over one year. Incidence rates assessed in a Dutch national study of diseases in general practice<sup>1</sup> suggest that 141 (0.8 per 1,000)

clinical hyperthyroid and 71 (0.4 per 1,000) clinical hypothyroid patients a year can be expected in the total population of 176,000 persons of the 80 GPs. A British incidence rate indicates that a regular GP in the United Kingdom detects 7 thyroid patients a year in a general practice population of averagely 2,500 persons<sup>2</sup>. On the basis of this rate, 493 new thyroid patients might be expected yearly in the population of 176,000. A partial explanation of the low incidence rates in the population sent to the DCC might be that GPs refer patients who show a clear picture of hyperthyroidism or hypothyroidism to a consultant (internist/endocrinologist) without previously requesting a TSH test themselves.

When the prediction instruments for clinical hyperthyroidism and clinical/subclinical hyperthyroidism are applied to the group of 668 patients, 14 (38%) of the 37 clinical hyperthyroid patients prove to have pretest probabilities of this diagnosis which are  $\leq 10\%$ ; regarding clinical/subclinical hyperthyroidism, equally low pretest probabilities are determined for 20 (34%) of the 58 patients with this diagnosis. Since the prediction instruments were developed using signs/symptoms, it can be surmised on the basis of these results that in general practice hyperthyroidism might not be diagnosed in a number of patients due to the circumstance that these show a not very typical picture of hyperthyroidism. This can also be the case with hypothyroid patients.

Hyperthyroid and hypothyroid patients with atypical signs/symptoms can remain undiagnosed in general practice (either under medical checkup or not) or can be treated for another diagnosis. It is also possible that such patients are referred to outpatient clinics because another disease is suspected on the basis of the patients' signs/symptoms. Atrial fibrillation (in hyperthyroidism) and a referral to a cardiologist as well as muscle complaints (in hypothyroidism) and a referral to a neurologist for instance come to mind.

To investigate the last-mentioned possibility, it was decided to retrospectively query the data base (case finding) of the Clinical Chemistry Laboratory of the University Hospital Maastricht regarding the period of 1 September 1993 until 1 July 1994 for patients who firstly belonged to the patient populations of the GPs participating in the prediction instrument development study and who secondly were diagnosed with hyperthyroidism or hypothyroidism by consultants in inpatient or outpatient clinics on the basis of abnormal TSH and Free Thyroxine (FT<sub>4</sub>) test outcomes. Additionally, it was decided to investigate by means of this retrospective study to what extent it would have been possible for a GP to diagnose hyperthyroidism or hypothyroidism on the basis of medical history and physical examination data prior to a patient's referral. The various clinical departments of the University Hospital Maastricht were asked permission for use of their inpatient and outpatient medical records in the study. Finally, it was studied as

well as possible how the diagnoses of hyperthyroidism or hypothyroidism were established. The following alternatives were taken into account:

- patients were referred by GPs to consultants because of a suspicion of hyperthyroidism or hypothyroidism without the GPs previously requesting a TSH test themselves;
- patients were referred to consultants for other reasons on the basis of signs/symptoms which were not recognized by GPs as being indicative of hyperthyroidism or hypothyroidism; the consultants in question did consider the possibility of these diagnoses on the basis of the signs/symptoms;
- in patients referred to consultants, hyperthyroidism or hypothyroidism was diagnosed by chance through a TSH test and a FT<sub>4</sub> test which were requested as part of a series of tests without suspecting these disorders.

### **Patients and methods**

The data base of the Clinical Chemistry Laboratory of the University Hospital Maastricht was queried for patients on whom thyroid function tests, i.e. a TSH test and a FT<sub>4</sub> test, were performed which indicated clinical hyperthyroidism or clinical hypothyroidism. Hyperthyroidism was diagnosed if the TSH test outcome was  $\leq 0.2$  mU/l and the FT<sub>4</sub> outcome was  $\geq 24$  pmol/l; hypothyroidism was diagnosed if the TSH outcome was  $\geq 5.0$  mU/l and the FT<sub>4</sub> outcome was  $\leq 10$  pmol/l. Additionally, the patients should meet the following criteria:

- be older than 14 years of age,
- not be pregnant,
- not be treated/checked up in connection with thyroid disease,
- be recently referred for clinical observation or outpatient examination by GPs participating in the prediction instrument development study,
- recently no thyroid function tests should have been requested for the patients by GPs, and
- the test requests should have been made during the period of 1 September 1993 until 1 July 1994.

Data of patients meeting the above criteria were obtained from the inpatient and/or outpatient clinics (medical records, encounter forms, letters of discharge, etc.). Since it concerned a retrospective study and data had not been acquired in a structured way, the prediction instruments could not be used to calculate the pretest probabilities of hyperthyroidism and hypothyroidism for the patients in question.

An internist was asked to estimate the patients' pretest probabilities of hyperthyroidism or hypothyroidism on the basis of the patient data, and whether he would have requested further diagnostics (a TSH test and a  $FT_4$  test) for these patients in order to confirm or exclude hyperthyroidism or hypothyroidism. The internist was informed of the facts that the patient group to be evaluated contained patients with either hyperthyroidism or hypothyroidism and that all available data (except the laboratory test outcomes) were presented.

## Results

Over the 10 months, 90 patients with clinical hyperthyroidism or clinical hypothyroidism were detected. Of 3 patients, not enough data were available to include them in the study; enough data were present about the remaining 87 patients.

Of this patient group, 46 patients were diagnosed with hyperthyroidism (10 males with an average age of 59 years and 36 females with an average age of 67 years), and 41 patients were diagnosed with hypothyroidism (10 males with an average age of 68 years and 31 females with an average age of 65 years).

Of the 87 patients, 55 patients were referred by GPs to outpatient clinics (38 patients to an internist, 8 patients to a cardiologist, 4 patients to a neurologist, 3 patients to a dermatologist, 1 patient to an orthopaedic surgeon, and 1 patient to a Throat/Nose/Ear (TNE) consultant). The average age of these patients was 61 years. The remaining 32 patients were referred by GPs for admission to hospital (18 patients to Internal Medicine, 7 patients to Cardiology, 5 patients to Pulmonology, 1 patient to the Intensive Care Unit (ICU), and 1 patient to Dermatology). Their average age came to 72 years.

The consulted internist estimated that, judging from their signs/symptoms, 18 (47%) of the 38 patients referred to the outpatient Internal Medicine clinic had fairly high probabilities of hyperthyroidism or hypothyroidism. The internist certainly would have requested a TSH test for these patients. Low pretest probabilities of these disorders were estimated concerning 5 patients (13%). For these patients, the internist possibly would have requested a TSH test. Regarding the remaining 15 patients (40%), the possibility of hyperthyroidism or hypothyroidism was considered highly improbable on the basis of the signs/symptoms. For them, the internist would not have requested a TSH test.

A thyroid disorder was considered in case of for example combinations of complaints like excessive perspiration, heart palpitations, fatigue, and weight decrease; slow mobility, weight increase, and oedemata; and use of Amiodarone (Cordarone). The

possibility of hyperthyroidism or hypothyroidism was considered improbable in case of for example muscular pains related to the use of Simvastatine (Zocor) to treat hypercholesterolaemia; and in case of combinations of complaints like obesity and hypertension or loss of appetite and weight decrease related to complaints regarding the upper abdomen.

In the referral letters written by GPs when referring patients to the outpatient Internal Medicine clinic, hyperthyroidism or hypothyroidism was mentioned as a possible diagnosis regarding 8 (21%) of the 38 patients. With respect to 27 patients (71%), no indication was found in the referral letters of whether hyperthyroidism or hypothyroidism was considered. The referral letters concerning the remaining 3 patients were absent.

Regarding the 8 patients who were referred to the outpatient Cardiology clinic, the consulted internist thought that hyperthyroidism or hypothyroidism was probable in 3 and improbable in the other 5 patients. These 5 patients had complaints which for instance related to angina pectoris with or without an increased cholesterol level or to a pacemaker checkup in a struma patient who showed no signs/symptoms indicative of thyroid disease.

In respect of the 9 patients who were referred to the outpatient Neurology, Dermatology, Orthopaedics, and TNE clinics, the consulted internist thought that hyperthyroidism or hypothyroidism was probable in 6 patients (67%) and improbable in the remaining 3 patients (33%). Regarding 3 patients who were referred to the outpatient Neurology clinic, the internist suspected hyperthyroidism in 1 patient suffering from a definite tremor in combination with fatigue and weight loss, and hypothyroidism in the other 2 patients of whom one suffered from muscular weakness combined with slow mobility and the other from loss of muscular strength combined with facial puffiness.

With respect to the 18 patients admitted to the inpatient Internal Medicine clinic, the consulted internist clearly thought of the possibility of hyperthyroidism or hypothyroidism in 7 patients (39%). Among these were 3 patients who were taking Amiodarone (Cordarone), which is indicative of possible thyroid disease, and 3 patients who suffered from atrial fibrillation in combination with a few other complaints indicative of possible hyperthyroidism. Regarding the remaining 11 patients (61%), the internist thought that hyperthyroidism or hypothyroidism was highly improbable. Among these were 2 patients with carcinoma of the colon; 1 patient who experienced relapses of collapsing and was incapable of taking care of himself; 1 patient who suffered from general malaise, relapses of pneumonia, and complaints when swallowing; 1 patient with mesenteric thrombosis; and 1 patient who had abdominal complaints and preterminal renal insufficiency.

Of the 14 patients admitted to the inpatient Cardiology, Pulmonology, Dermatology, and ICU clinics, the consulted internist suspected 4 patients (29%) of possibly having hyperthyroidism or hypothyroidism. The probability of these diagnoses was considered doubtful in another 2 patients (14%) and highly improbable in the remaining 8 patients (57%).

## Discussion

GPs in the Maastricht region send patients via the DCC to the University Hospital Maastricht to have further diagnostic tests performed by means of diagnostic hospital facilities. The DCC supports the GPs in their diagnostic behaviour, for instance, by stipulating work procedures about indications for diagnostic test requests and by giving periodic feedback about these requests. The feedback takes the form of notes about for example instances in which a GP does not abide by work procedures agreed upon or instances in which requests of diagnostic tests do not seem rational. This is for instance the case if a pretest probability of a disease is very low, or if a diagnostic test is chosen which has too little discriminative power regarding a particular disease. The aim is to have GPs request further diagnostics as rationally as possible. In this decision process, the work procedures and principles of medical decision making are important factors.

**Table 1.** Number of requested tests and number of abnormal test outcomes

Number of Tests per GP	Number of GPs		Number of tests per group		Number of abnormal outcomes		Abnormal outcomes per number of tests
	Abs.	%	Abs.	%	Abs.	%	
1 - 5	27	34	104	15	7	11	7 %
6 - 10	36	45	298	45	33	50	11 %
11 - 15	10	12	125	19	12	18	10 %
> 15	7	9	141	21	14	21	10 %
Total	80		668		66		

Data concerning the application of these principles of medical decision making are taken from literature<sup>3-5</sup>. Examples are for instance data which enable an estimation of a pretest probability regarding a disease. This type of data often is known with respect to hospital but not in relation with general practice. Exactly for the primary care environment, prediction instruments for thyroid disorders were created with which pretest probabilities can be calculated using signs/symptoms.

The development of these instruments was based on 668 patients. Concerning these patients, the GPs considered the possibility of hyperthyroidism or hypothyroidism on the basis of signs/symptoms or requested a TSH test for instance on a patient's wish. Table 1 provides information about the number of requested TSH tests by the 80 GPs and the number of abnormal test outcomes (clinical or subclinical hyperthyroidism or clinical hypothyroidism) cumulatively. The number of requested tests varies from GP to GP between 1 and > 15. The percentage of abnormal test outcomes is more or less the same for groups of GPs who request TSH tests in divergent numbers. The number of clinical or subclinical hyperthyroid and clinical hypothyroid patients (66) is small compared to the number of expected diagnoses based on incidence and prevalence rates mentioned in literature<sup>1,2,6,7</sup>. Consequently, the question was raised whether the incidence and prevalence rates of clinical or subclinical hyperthyroidism or hypothyroidism, for whatever reason, are lower in the Maastricht region than elsewhere, or whether these diagnoses were missed by GPs, or whether patients having signs/symptoms on the basis of which these disorders can be suspected are referred to outpatient or inpatient clinics. Since these questions emerged in the course of the mentioned research, it was not possible to study them prospectively. In a retrospective study, it was investigated how many patients who were referred to outpatient or inpatient clinics in the period of 1 September 1993 until 1 July 1994 were diagnosed with hyperthyroidism or hypothyroidism on the basis of performed diagnostics (a TSH test and a FT<sub>4</sub> test). This proved to be the case with 90 patients.

To get an as good as possible idea about the alternatives which lie at the basis of the diagnoses of hyperthyroidism and hypothyroidism and are mentioned in this chapter's Introduction, all inpatient and outpatient data of 87 patients were studied. With respect to a few patients only, it proved possible to obtain an indication of the degree of an attending consultant's suspicion of such a diagnosis prior to the TSH test and the FT<sub>4</sub> test. Usually, consultants did not mention a suspicion to this effect in their correspondence about these patients. The internist who studied the data retrospectively tried to indicate as good as possible on the basis of the inpatient and outpatient data to what extent he thought a suspicion of hyperthyroidism or hypothyroidism was justified using medical history and physical examination data and regarding which patients he would have requested a TSH test and a FT<sub>4</sub> test on the basis of such suspicions in order to confirm or exclude these diagnoses. The internist estimated that 46% of the patients referred to the outpatient Internal Medicine or Cardiology clinics had sufficiently high pretest probabilities of hyperthyroidism or hypothyroidism to justify the TSH and FT<sub>4</sub> test requests. This also

proved to be the case with 67% of the patients referred to the other outpatient clinics. The probability of hyperthyroidism or hypothyroidism, furthermore, seemed to be high in 39% of the patients admitted to the inpatient Internal Medicine clinic. This also proved to be the case with 29% of the patients admitted to the other inpatient clinics.

All this means that a majority of the patients referred by GPs to the inpatient and outpatient clinics probably was not suspected of having a thyroid disorder on the basis of medical history and physical examination data. A thyroid disorder was diagnosed by consultants using further diagnostics within which the TSH test and the FT<sub>4</sub> test were performed as part of a series of laboratory tests. That the diagnosis of a thyroid disorder was fortuitous for some patients can be concluded from the fact that the abnormal laboratory test outcomes were not mentioned in the letters of discharge written to GPs.

Since the study was carried out retrospectively, the results should be interpreted with caution. There is a possibility that a thyroid disorder was considered with respect to some of the patients for whom the consulted internist estimated this not to be justified, but that related data were not included in the medical records. The data, by the way, seem to indicate that hyperthyroidism or hypothyroidism was diagnosed on unsuspected grounds in a higher percentage of the studied inpatients than in patients referred to outpatient clinics, although the difference is not significant. The higher age of the clinically admitted patients and the presence of (more) comorbidity in this group of patients possibly will have accounted for this.

The following remarks can be made with respect to the GP situation:

- GPs considered a diagnosis of hyperthyroidism or hypothyroidism in 21% of the patients referred to the outpatient Internal Medicine clinic. The consulted internist also was of the opinion that all of these patients had sufficiently high pretest probabilities of hyperthyroidism or hypothyroidism to justify the TSH and FT<sub>4</sub> test requests. GPs should have requested these diagnostic tests themselves.
- Regarding another 27% of the patients who were referred to the outpatient Internal Medicine clinic and who were estimated by the consulted internist to have an increased probability of hyperthyroidism or hypothyroidism, the latter could not be concluded from the GPs' referral letters. It is not clear what the reasons were for these differences in judgement. However, it can be noted that the consulted internist was informed of the eventual diagnoses.
- GPs did not consider the diagnosis of hyperthyroidism or hypothyroidism in the patients referred to the other outpatient clinics. If, however, they would have considered these diagnoses, the GPs should have referred the patients probably to



the outpatient Internal Medicine clinic.

- It is not clear to what extent GPs considered the diagnoses of hyperthyroidism or hypothyroidism with respect to the admitted patients diagnosed with these thyroid disorders. Since another and usually clearly manifest disease was the reason for admission, a GP probably will not have considered a thyroid disorder. To what extent consultants did consider a thyroid pathology can not be ascertained from the clinical data. The consulted internist estimated that 39% of the patients admitted to the inpatient Internal Medicine clinic had sufficiently high pretest probabilities of hyperthyroidism or hypothyroidism to justify the thyroid function test requests.

In conclusion, it can be stated that the retrospective study has offered too few possibilities to definitely establish the reasons why hyperthyroid or hypothyroid patients diagnosed in outpatient or inpatient clinics were not diagnosed as such previously by the GPs. The clinical pictures of a number of patients will have been insufficiently clear to the GPs to indicate a suspicion of hyperthyroidism or hypothyroidism. A prospective study can provide more definite answers. In such a study, the developed prediction instruments can be of use.

## References

1. van der Velden J, de Bakker DH, Claessens AAMC, Schellevis FG. Een nationale studie naar ziekten en verrichtingen in de huisartspraktijk - Basisrapport: Morbiditeit in de huisartspraktijk. Utrecht: Nederlands Instituut Voor onderzoek van de EersteLijnsgezondheidszorg (NIVEL), 1991.
2. Hart IR, Newton RW. The new medicine: An integrated system of study - Volume 2: Endocrinology. Lancaster: MTP Press Limited, 1983.
3. Bouter LM, van Dongen MCJM. Epidemiologisch onderzoek - Opzet en interpretatie. Houten: Bohn, Stafleu & van Loghem, 1991.
4. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology - The essentials. Baltimore: Williams & Wilkens, 1988.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: A basic science for clinical medicine. Boston: Little, Brown and Company, 1991.
6. Wiersinga WM, Krenning EP. Schilddklierziekten. Alphen aan den Rijn: Samsom Stafleu, 1989.
7. van Es A, Geerdink RA. Schilddklierziekten. Utrecht: Bunge, 1992.

## Acknowledgements

Special thanks are expressed to the Departments of Clinical Chemistry, Internal Medicine (and its clinics General Internal Medicine, Endocrinology, Geriatrics, Oncology, Nephrology, Rheumatology, and Gastro-enterology), Cardiology, Dermatology, Throat/Nose/Ear, and Neurology of the University Hospital Maastricht for their cooperation in the retrospective study. The panel of experts and the consulted internist are thanked for their valuable work in evaluating medical records of patients.

## **Chapter 10**

### **Conclusions and general discussion**

## Introduction

In this chapter, principal research findings are reviewed, consequences of the research and findings in literature are discussed, and recommendations for future research are made.

## Principal findings

The studies described in this thesis dealt with primary care decision support. Research was performed with respect to Information Technology (IT) and its application to medicine. Furthermore, prediction instruments were built which can help General Practitioners (GPs) in assessing the necessity of ordering a Thyroid-Stimulating Hormone (TSH) test.

The Phased Development Methodology (PDM) in combination with a system architecture consisting of a data base, a model base, a user interface and system management software proved to be an adequate environment for medical software development. Decision Support Systems (DSSs) which offer active and/or passive decision support were integrated with a GP information system. Both the GP information system HIOS and the DSS HIOS+ were positively evaluated in laboratory and field studies.

The principal questions of the IT research formulated in chapter 1 could be answered:

- *Can the quality of registered diagnoses and the completeness of medical records be enhanced by DSSs?*
- It was concluded that active support for checking whether the criteria for diagnoses are indeed fulfilled led to a better quality of registered diagnoses and more complete medical records. This kind of active support in particular resulted in an increase of GPs' diagnostic correctness (misdiagnoses were prevented), in a higher diagnostic confidence of GPs, and in a more complete data capture. Passive support proved to be less effective. Especially in cases which GPs thought were not too difficult, the supply of additional information was considered superfluous. Yet, a considerable number of errors remained unrecognized by the GPs.
- *With which type of decision support - active or passive - provided by DSSs are GPs most satisfied?*
- Active support regarding the interpretation of diagnostic criteria was positively evaluated by its users in both a laboratory study and a field study. In the

multigroup laboratory study, it was determined that the group of GPs given this kind of active support was significantly more satisfied with the system than the GPs in the groups which either were not supported or were given passive support. The GPs in the field study were even more satisfied. The observed user satisfaction levels definitely justify the incorporation of active decision support in primary care DSSs.

- *What is the GPs' level of knowledge regarding the medical domain of thyroid disorders when compared to the knowledge stored in the models of a DSS?*
- GPs' medical (diagnostic) knowledge of thyroid disorders was tested by a DSS on the basis of diagnostic items entered by them. It was found that the overall knowledge of GPs about thyroid disorders was reasonably good. GPs with the most years of work experience scored best. According to the GPs, the DSS can be used in education and refresher courses.

The principal questions of the research into thyroid disorder decision making formulated in chapter 1 also could be answered:

- *Can prediction instruments based on signs and symptoms be developed which indicate whether a TSH test is adequate?*
- This question could be answered positively on the basis of the results of a prospective study. With the developed instruments, a GP can more effectively determine whether a TSH test request is adequate. The percentage of false positive predictions, furthermore, can be decreased. The reduction in test requests, however, is dependent on the rate of false negative predictions which still is considered acceptable by a GP.
- *How many patients with thyroid disease for whom GPs do not request a TSH test are detected in hospital; where are they detected; and why are they not detected in primary care?*
- In a retrospective study, 46 clinical hyperthyroid and 41 clinical hypothyroid patients were detected in inpatient and outpatient clinics over the same 10-month period as the initial data acquisition period of the prospective study. The number of hyperthyroid patients suggests that for each clinical hyperthyroid patient detected in primary care, a clinical hyperthyroid patient was detected in hospital. With respect to hypothyroidism, the detection rate even was 5 times higher in

hospital than in general practice. A consulted internist indicated that he would have definitely requested a TSH test in 44% of the cases, that he would have been in doubt about whether or not to request a TSH test in 8% of the cases, and that he would not have requested a TSH test in the remaining 48% of the cases. It has not become clear from the retrospective study with its limited possibilities why the GPs did not consider the diagnoses of hyperthyroidism or hypothyroidism in these patients before referral. Further investigations are necessary.

## **Consequences of the research and findings in literature**

### *Development methodologies*

In literature, many software development methodologies (e.g. the Information Systems work and Analysis of Change (ISAC) Method, the Business Systems Planning (BSP) Method, and the System Development Method (SDM)) are described<sup>1,2</sup>. Most of these methodologies combine several techniques, e.g. the use of Data Flow Diagrams (DFDs) and prototyping. In the PDM software environment, use was made of several of the available methodologies with included techniques, like prototyping, the top-down and/or bottom-up technique, and iterative design, and user involvement was emphasized. Users were involved in the determination of system requirements and implementation priorities, and participated in the evaluation of several prototypes.

It was experienced that evolutionary prototyping<sup>3</sup> not only shortened the development process of medical and DSS applications, but also that it increased the quality of systems in terms of the coverage of system requirements. Discrepancies between the developer's and the users' way of thinking could also be detected faster, leading to a reduction in software development costs. The usefulness and the attainability of the specifications regarding the patient registration module in the 1985 reference guide of the Dutch College of General Practitioners were confirmed in the created developments. Specifications of the medical module stipulated in the research were provided to the College and were even used in its 1990 and 1995 reference guides.

In literature, however, also pitfalls in prototyping have been reported. These concern instances in which: (i) alterations in preliminary specifications are easily made and are not restricted, (ii) iterations of prototype versions abound because sound reviews of the prototypes are not performed, milestones are not established, and a deadline for delivery is not set, (iii) users do not have the required knowledge to participate in the prototyping process, and (iv) users do not provide specifications which are acceptable to all users of a system. The mentioned problems can be put down to bad management. In

the development of HIOS and HIOS+, it was made sure that these problems were avoided.

#### *The HIOS+ system*

In HIOS+, DSSs are integrated with the GP information system HIOS. The system was used in daily practice, and it was there that the advantages of integration disclosed themselves: data only had to be entered once into the data base and subsequently could be used by the different DSSs within HIOS+; and each time a new DSS was developed, it could easily be integrated in HIOS+. Nowadays, a number of commercial GP information systems are on the market. Their functionalities in terms of decision support, however, are rather limited. It is expected that DSS packages will become available in the near future. Since these packages in the current situation will have to be used as stand-alone systems, suppliers of GP information systems should deliver open systems and use standardized terminology to enable integration.

#### *Acceptation by GPs*

All software needs to meet the functional and technical requirements specified by the GPs involved<sup>4,5</sup>. Users will be satisfied if their wishes are taken into account in the technical realization of systems. Only if the user satisfaction is high<sup>6,7</sup>, the introduction of these systems in general practice can be successful.

Adapted Bailey questionnaires were used to assess aspects of user satisfaction with HIOS+ and its DSSs. When used at regular intervals, the Bailey questionnaire<sup>6,7</sup> is an instrument with which the satisfaction of users with information systems can be determined and changes in user satisfaction can be detected. As a management instrument, the Bailey questionnaire can be applied in general to evaluate performed activities and their impact on user satisfaction scores.

In the laboratory study, user satisfaction was assessed only once. Users were taught how to work with the system at the day the study was held. The idea behind the study was to identify possible areas for improvement with respect to the functionality. It was not known whether the Bailey questionnaire is a valid instrument for this task. However, known limitations of the test site indeed were corroborated on the basis of aspect scores. This at least is a positive indication of the validity of the Bailey questionnaire for this task. A few other limitations, e.g. the completeness and precision of some of the criteria used in HIOS+, also became clear.

The laboratory study showed that GPs were moderately satisfied with the

integrated system. After the field study, in which GPs worked with the system for three months in daily practice, nevertheless, the GPs indicated to be quite satisfied with HIOS+. This more positive result was expected since the field-study GPs better learned to use the system's functionalities after longer practice.

### *Most effective decision support*

It was established that unsolicited passive support did not lead to a major improvement in the quality of registered diagnoses. Active decision support, which has a watchdog function, on the other hand resulted in such an improvement and was appreciated for its achievements by the GPs.

Usually GPs were confident of their diagnoses and did not bother to check them by reading additional information. It can be expected that GPs will not ask for diagnostic support if they are sure of their diagnoses. On the basis of the laboratory research, it may be added that in such cases GPs probably will not read additional information either. In the laboratory study, about 8% of the diagnostic conclusions of the GPs were wrong, even though the GPs were certain about their diagnoses. Active support will lead to a higher diagnostic accuracy and concomitantly to a more complete medical record.

In order for a DSS to perform its watchdog function, a more complete data entry is necessary. It appeared that GPs did not enter many data spontaneously, and, therefore, additional data had to be asked for by the system. A more complete data entry also can be obtained by using dynamic entry lists. Advantages of both approaches are, e.g., a more standardized data acquisition and a higher quality of registration. A disadvantage is the more restricted way of working.

It was experienced that the free-text criteria of the ICHPPC-II-Defined often were not structured or detailed enough for formalization. If the original free-text terms would have been used in the DSS, the chances would have been that GPs would use other terms during patient encounter registration, which would have constituted a problem for the DSS. A structured data entry procedure alleviates this problem. In order for the DSS to be able to check each diagnosis, many data have to be entered in its data base. Reliable data acquisition methods for bulks of data still have to be found.

### *GPs' knowledge of thyroid disorders*

It can be concluded that the knowledge implemented in the models about thyroid disorders is known to GPs. GPs, however, have to establish diagnoses in daily life on the basis of patients' clinical pictures which can not always be easily discriminated from

clinical pictures of other diseases. Comorbidity, the effects of used medication, and the incompleteness of the information provided by patients during encounters can also complicate the diagnosis of thyroid disorders.

#### *Guidelines for thyroid disease*

In The Netherlands, the Dutch College of General Practitioners issued many guidelines for GPs. One of these concerns thyroid disorders<sup>8</sup>. In this guideline, prevalence rates of hyperthyroidism and hypothyroidism are stated, drug policies are presented, and clinical pictures and preferred laboratory diagnostics are described. It is also discussed under what circumstances patients should be treated by GPs and when patients should be referred to hospital. No predictive values of signs/symptoms are indicated because these were unknown. In general, the College's guidelines are a first step in stimulating a rational diagnostic behaviour of medical professionals.

Recently, the National Sickfund Board published a book with diagnostic guidelines<sup>9</sup>. The first chapter deals with the use of epidemiologic concepts in the interpretation of diagnostic tests. However, the principles of decision analysis are not extensively discussed and are not used in the other chapters of the book. The chapter on hormones and metabolism deals with thyroid disorders, but these diseases are only very generally presented without any reference to the use of decision-making principles in their diagnosis. At present, this book does not stimulate a more structured use of decision-making principles by GPs.

In Maastricht, twice a year personal feedback is given to GPs regarding their diagnostic request behaviour in a randomly chosen month. This feedback is based on work procedures, principles of medical decision making, and data provided by GPs on request forms.

#### *Prediction instruments*

Instruments were developed to select probable thyroid patients from those for whom a GP intends to request a TSH test. The prediction instruments described in this thesis depend on the identification of patients by GPs: patients who are not identified by GPs can not be analysed by the prediction instruments. However, it is nearly impossible to test an entire patient-encounter population due to the bulk of data which would be required. If also instruments would be available for other diseases which would have to be evaluated in general practice, data entry certainly would become a serious problem to GPs.



The prediction instruments can reduce the number of unnecessary test requests. However, the reduction which can be achieved if false negatives are not acceptable is limited. At a sensitivity of 100%, the specificity is about 3% for the instrument for clinical/subclinical hyperthyroidism, 24% for the instrument for clinical hyperthyroidism, and 13% for the instrument for clinical hypothyroidism. A sensitivity of less than 100% results in higher reductions in the number of test requests. In this case, though, GPs need to reach consensus on a percentage of false negative predictions which is still acceptable. The extra costs for the community which are attended with patients missed and the patients' quality of life also have to be considered. In the Maastricht region, feedback has been given for more than a decade. In a different region where less or no feedback is given, the use of prediction instruments may result in even considerably higher reductions in test requests.

The prediction instruments were developed using stepwise logistic regression. Prediction instruments based on signs/symptoms can also be created for other diseases. Decision trees and the use of Bayes' theorem are examples of alternative methods which can be used to perform this function.

The probability thresholds used in the prediction instrument study were low due to the fact that they were based on a population of patients who were more similar than a general practice population. These probability thresholds, hence, do not equal the pretest probabilities applied by GPs in their decision process. It can be deduced from the data that GPs apply a pretest probability of 10% for hyperthyroidism and of 8% for hypothyroidism to general practice populations.

The specificity of the prediction instruments could be determined with a patient group of a total of 436 patients during a one-month pilot study held in a health centre. All patients visiting the health centre were asked to participate. Use of the prediction instruments, in which the  $\beta_0$  had been adjusted for the general population, resulted in a TSH test request for one patient only. Over the same period, however, the GPs requested a TSH test for four patients. The patient for whom the prediction instruments advised a TSH test was a false positive. The four TSH tests requested by the GPs, though, all had outcomes indicating euthyroidism. The impression exists that a considerable request reduction can be achieved in general practices, provided that the sensitivities of the instruments remain the same. It will be interesting to investigate whether the prediction instruments can be used on GP-referred patients diagnosed with hyperthyroidism or hypothyroidism in hospital to determine if these diagnoses should have been detected by the GPs prior to referral.

### *Present request situation*

GPs request tests using forms with different layouts, in which many or a limited number of tests are printed on a form, tests are categorized on the basis of laboratory departments where they are performed, and tests are categorized in a problem-oriented way based on signs, symptoms, and diseases. A layout can influence the test request behaviour. The use of a problem-oriented layout, in which clusters of tests are mentioned belonging to a number of problems, can improve test adequacy. In this approach, however, patients' clinical pictures are not taken into account in that pretest probabilities of diseases are not estimated. The use of guidelines and work procedures based on medical decision-making principles on the other hand can influence the test request behaviour much more.

Until recently, software for the support of the feedback process was independently developed in the Maastricht<sup>10</sup> and Nijmegen<sup>11</sup> regions. The Nijmegen system can establish differential diagnoses. In that case, sometimes rare diseases are listed which may consequently lead to an increase in the number of requested tests.

It is expected that the use of a DSS in giving feedback will provide even more advantages than retrospective (manual) feedback, namely:

- additionally avoid unnecessary requests and costs<sup>11-19</sup>;
- provide a GP with immediate active support<sup>20-23</sup>; and
- enable a reliable and fast exchange of data<sup>15,19,24</sup> between requesting GPs and laboratories or diagnostic centres.

Therefore, a nationwide project has been initiated in which several universities, hospitals, diagnostic centres, and GP information system suppliers, like Euroned Systems and PharmaPartners B.V., participate with the aim to develop such a laboratory test-ordering DSS, integrate it with existing GP information systems, and perform a multicentre evaluation study.

### *Identification of patients with a thyroid disease*

It was found that many patients were diagnosed with thyroid disorders in different clinics of the hospital. However, only 8 of the 87 thyroid patients (9%) diagnosed in hospital and having a typical picture were immediately referred to the Endocrinology outpatient clinic. For these patients, a consulted internist indicated that a TSH test request would have been adequate. It is not clear why GPs did not request a TSH test before referral. Signs and symptoms indicative of hyperthyroidism or hypothyroidism possibly were not recognized. Further investigations need to be performed.

### *Differences in incidence rates*

Incidence rates found in literature<sup>25</sup> suggest higher numbers of clinical hyperthyroid (0.8 per 1,000) and clinical hypothyroid (0.4 per 1,000) patients than were found in the prospective and retrospective studies together (0.5 per 1,000 for hyperthyroidism and 0.3 per 1,000 for hypothyroidism). This might mean that thyroid patients remained undetected in the patient population, or that incidence rates are different for the Maastricht region. The latter, however, is not likely. Possible explanations may also be that the general incidence rates were based on population survey results, or that many more TSH tests were requested for patients with atypical signs or symptoms or in the line of checkups by GPs in incidence rate assessment than were requested in the Maastricht region. The latter would undoubtedly lead to more detections of thyroid patients.

Contrary to indications in literature<sup>26-29</sup> that many patients with a thyroid disease are referred to hospital Psychiatry clinics (due to misdiagnoses by GPs), in the retrospective study only one thyroid patient was detected in the Psychiatry (outpatient or inpatient) clinics of the University Hospital Maastricht. The psychiatric facility Vijverdal was also taken into account in the case finding. The impression exists that GPs were aware that depression may mask a thyroid disorder and requested TSH tests for depressive patients, which resulted in several detections.

### **Recommendations for future research**

On the basis of the research described in this thesis, recommendations can be made for future research. These proposals relate to medical or technological issues. In future, research can be carried out:

- To further evaluate the prediction instruments in general practices.
- To prospectively study patients who are referred to hospital for other reasons than thyroid disorders and for whom abnormal TSH and FT<sub>4</sub> outcomes are determined in hospital but for whom no TSH tests are previously requested by GPs. In a prospective study, up-to-date information can be obtained from GPs. As was mentioned earlier, the prediction instruments can be used in this investigation.
- To further investigate data entry methods which can be applied to acquire data in the form of standardized terminology required for use within DSSs.
- To identify other medical domains in which GPs need support on the basis of epidemiological data.

## References

1. Sommerville I. Software engineering. New York: Addison-Wesley Publishing Company Inc, 1992.
2. Davis GB, Olson MH. Management Information Systems: Conceptual Foundations, Structure, and Development. New York: McGraw-Hill 1985.
3. Budde R. Prototyping: an approach to evolutionary system development. Berlin: Springer Verlag Heidelberg/Berlin, 1992.
4. Gillies J, Barry D, Crane J, Jones D, MacLennan L, Pearce N, Reid J, Toop L. A community trial of a written self management plan for children with asthma. *N Z Med J* 1996;109(1015):30-33.
5. Grogan S, Conner M, Willits D, Norman P. Development of a questionnaire to measure patients' satisfaction with general practitioners' services. *Br J Gen Pract* 1995;45(399):525-529.
6. Bailey JE. Development of an instrument for the management of computer user attitudes in hospitals. *Meth Inf Med* 1990;29(1):51-56.
7. Bailey JE, Pearson SW. Development of a tool for measuring and analyzing computer user satisfaction. *Management Science* 1983;29(5):530-545.
8. Pop V, Boer AM, Winants Y, Bommel van-Uittenhout AM, De voogd van der Straaten I, Smeele I, Felix-Schollaart B. NHG-standaard functiestoornissen van de Schildklier. *Huisarts en Wet* 1993;36(4):143-149.
9. Ziekenfondsraad. Diagnostisch Kompas. Utrecht: Boekhoven-Bosch:1997.
10. Boran G, O'Moore R, Grimson W, Peters M, Hasman A, Groth T, van Merode F. A new clinical laboratory information system architecture from the Openlabs project offering advanced services for laboratory staff and users. *Clinica Chimica Acta* 1996;248:19-30.
11. Oosterhuis WP, van den Bosch NJ, van den Hoogen HJ, van de Calseijde JF. Ondersteuning van de huisarts bij het aanvragen van laboratoriumdiagnostiek. Veldhoven: VMBI/MIC, 1994.
12. Lyon AW, Greenway DC, Hindmarsh JT. A strategy to promote rational clinical chemistry test utilization. *Am J Clin Pathol* 1995;103(6):718-724.
13. Connelly DP, Bennett ST. Expert systems and the clinical laboratory information system. *Clin Lab Med* 1991;11(1):135-51.
14. Rosser WW, McDowell I, Newell C. Use of reminders for preventive procedures in family medicine. *Can Med Assoc J* 1991;145(7):807-814.
15. Foy JL, Eastman RC, Nealon RC, Bowen PM, Pengelly ML, Drass JA, Dorworth TE, Pucino F. Automated therapeutic drug monitoring in an ambulatory care endocrine clinic. *Ann Pharmacother* 1992;26(5):675-678.
16. Friedman BA. The potential role of physicians in the management of hospital information systems. *Clin Lab Med* 1990;10(1):239-250.
17. Frank OR. Registers, recalls and reminders. *Aust Fam Physician* 1997;26(1):42-47.
18. Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *J Am Med Inform Assoc* 1996;3(6):399-409.
19. Turner BJ, Day SC, Borenstein B. A controlled trial to improve delivery of preventive care: physician or patient reminders? *J Gen Intern Med* 1989;4(5):403-409.
20. Nilasena DS, Lincoln MJ. A computer-generated reminder system improves physician compliance with diabetes preventive care guidelines. *Proc Annu Symp Comput Appl Med Care* 1995;640-645.
21. Szilagyi PG, Rodewald LE, Savageau J, Yoos L, Doane C. Improving influenza vaccination rates in children with asthma: a test of a computerized reminder system and an analysis of factors predicting vaccination compliance. *Pediatrics* 1992;90(6):871-875.
22. Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD. A computer-based monitoring system for follow-up of elevated blood pressure. *Med Care* 1983;21(4):400-409.
23. McDonald CJ, Murray R, Jeris D, Bhargava B, Seeger J, Blevins L. A computer-based record and clinical monitoring system for ambulatory care. *Am J Public Health* 1977;67(3):240-245.
24. Lobach DF. Electronically distributed, computer-generated, individualized feedback enhances the use of a computerized practice guideline. *Proc AMIA Annu Fall Symp* 1996; 493-497.
25. van der Velden J, de Bakker DH, Claessens AAMC, Schellevis FG. Een nationale studie naar

ziekten en verrichtingen in de huisartspraktijk - Basisrapport: Morbiditeit in de huisartspraktijk. Utrecht: Nederlands Instituut Voor onderzoek van de EersteLijnsgezondheidszorg (NIVEL), August 1991.

26. Howland RH. General Health, Health Care Utilization, and Medical Comorbidity in Dysthymia. *Int J Psychiat Med* 1993;23(3):211-238.
27. Fava M, Labbate LA, Abraham ME, Rosenbaum JF. Hypothyroidism and Hyperthyroidism in Major Depression Revisited. *J Clin Psychiat* 1995;56(5):186-192.
28. Ryan WG, Roddam RF, Grizzle WE. Thyroid Function Screening in Newly Admitted Psychiatric Inpatients. *Ann Clin Psychiat* 1994;6(1):7-12.
29. Enns M, Ross C, Clark P. Thyroid Screening Tests in Psychiatric Inpatients. *Gen Hosp Psychiat* 1992;14(5):334-339.

## Chapter 11

### Summary

The purpose of this thesis is to give a scientific discourse of the research into the field of diagnostic decision support for General Practitioners (GPs). Information Technology (IT) in primary care decision support and thyroid disorder decision making in primary care are the two main areas of research dealt with. In detail, the thesis reports of the research into:

- the feasibility of computer-based decision support in primary care, i.e. the application of IT to support the decision process of GPs, and
- the structuring of medical knowledge concerning the medical domain of thyroid disorders for decision support purposes, i.e. the development of (computer-based) prediction instruments for thyroid disorders with which GPs can determine Thyroid-Stimulating Hormone (TSH) test adequacy in patient cases.

Background information about the research, the principal research questions, and the structure of the thesis are described in **chapter 1**. The research was actuated by a survey among GPs which theoretically had shown that GPs had a need for decision support by means of Decision Support Systems (DSSs), that both active and passive support were required in many instances, and that the medical domain of thyroid disorders was a problem area in which decision support was welcome. The principal research questions of the IT research were:

- *What is the state of the art of GP information systems and DSSs?*
- *Can the quality of registered diagnoses and the completeness of medical records be enhanced by DSSs?*
- *With which type of decision support - active or passive - provided by DSSs are GPs most satisfied?*
- *What is the GPs' level of knowledge regarding the medical domain of thyroid disorders when compared to the knowledge stored in the models of a DSS?*

The principal research questions of the thyroid disorder knowledge research were:

- *Can prediction instruments based on signs and symptoms be developed which indicate whether a TSH test is adequate?*
- *How many patients with thyroid disease for whom GPs do not request a TSH test are detected in hospital; where are they detected; and why are they not detected in primary care?*

Whereas chapters 1, 2, and 10 describe both research areas, chapters 3 and 5 through 7 deal with the IT research and chapters 4, 8, and 9 cover the thyroid disorder knowledge research.

**Chapter 2** contains the results of a literature review, and is composed of three parts. The GP Information Systems part describes findings in literature relating to the evolution of primary care systems, to the development, (dis)advantages, and requirements for a proper function of primary care patient information systems, and to the present situation of computer and other tool usage by GPs in The Netherlands. The Decision Support for GPs part deals with findings in literature relating to the evolution of system types, DSS development and implementation, DSS evaluation, and the medical impact of DSSs. The Medical Knowledge of Thyroid Disorders part deals with findings in literature relating to the test ordering of GPs and to past and present tools for diagnosing thyroid disorders.

In **chapter 3**, the fully integrated GP Information and Research System HIOS and DSS HIOS+ are described. The systems were developed using the Phased Development Methodology (PDM). HIOS consists of six modules, has all the characteristics of a medical registration system to support GPs in daily practice activities, and offers research facilities. It was implemented at a general practice and a health centre, and evaluated positively. HIOS+ consists of HIOS and seven decision support modules/DSSs. It assists GPs in (diagnostic) decision making by presenting textbook information, generating advice and warnings, creating models, generating reports, and evaluating knowledge for educational purposes. All evaluations of HIOS+ modules by GPs had positive results. The full integration of a medical registration system with DSSs, and the use of a model base separately from the software are advantageous. It was concluded that HIOS and HIOS+ offer diagnostic and disease management assistance, and can be used to screen and monitor patient populations.

**Chapter 4** describes the design of the prospective patient-oriented study of thyroid disorder prediction instruments. Data were acquired using general procedures for TSH and Free Thyroxine (FT<sub>4</sub>) test ordering by GPs and test outcome interpretation, study-specific procedures for inclusion criteria application and questionnaire handling, and created and validated patient and GP questionnaires and the Diagnostic Coordinating Centre (DCC) document. From 1 September 1993 until 1 July 1994, 80 Maastricht region GPs requested a TSH test for 668 new patients (37 clinical hyperthyroid, 21 subclinical hyperthyroid, 8 clinical hypothyroid, 37 subclinical hypothyroid, and 565 euthyroid patients). Their data were analysed concerning data quality in terms of completeness and correctness. The data entry and data base quality was monitored with the registration



system THYPRED. In data analysis for instrument creation, the data were analysed univariately, bivariately, and multivariately (factor and stepwise logistic regression analyses). In instrument evaluation, the Grouped Jackknife Method was used.

In **chapter 5**, an IT experiment is described which concerned the usefulness of the DSS Formal Interpreter (FI) in the validation of International Classification of Primary Care (ICPC) codes during data registration using free-text International Classification of Health Problems in Primary Care (ICHPPC)-II-Defined criteria. Computer-processable queries were created with the DSS Free-text Formalizer for 101 ICPC codes, and were validated by an independent national GP panel. In the experiment, three equally large GP groups, given no, FI's passive decision support (criteria presentation), and FI's active decision support (advice and warnings based on query processing with medical record data) respectively, were presented case descriptions for which diagnoses and justifying data had to be registered. These data were validated by a GP and an independent GP panel, and analysed, also regarding statistically significant group differences (the Mann-Whitney test), by the researcher with respect to FI's impact. It was concluded that active decision support results in a higher diagnostic correctness and confidence, and a more complete medical record; that free-text criteria can not be straightforwardly translated into queries; and that the Arden Syntax for Medical Logic Modules is flexible in query creation.

**Chapter 6** deals with an IT experiment of assessing user satisfaction with FI. In a laboratory test, three equally large GP groups, given no decision support (group I), passive support (group II), and active support (group III) respectively, were asked to enter case description data. User satisfaction was assessed with an adapted Bailey questionnaire. 'Slightly satisfactory' average and normalized group scores were computed regarding functionality, performance, flexibility, the users' feeling of participation, and FI's integration with HIOS. Group III GPs were more satisfied than the other GPs; group II GPs were least satisfied. A statistically significant difference (determined with the two sample t test) existed between groups II and III, but not between groups I and III; FI's questioning of group III GPs presumably was not a problem. Free-discussion questions showed that the GPs considered active and passive support generally useful, would have liked to receive both types of support, and expected a variety of advantages from them. In a field test, FI was applied by GPs in a rural general practice for three months. 'Quite satisfactory' average and normalized scores were assessed using the adapted Bailey

questionnaire. It was concluded from the experiment that FI's very satisfactory test results gave good hope that medical DSSs in future would be used by GPs.

In chapter 7, an IT experiment is described which concerned the DSS Evaluator (EV) and its function of testing medical knowledge and giving decision support in case of knowledge gaps. In the experiment, 17 GPs applied EV to assess their knowledge of thyroid disorders using case descriptions. A GP had to specify relations between a (sub)diagnosis and gender or age, and to select knowledge items (e.g. signs/symptoms, and laboratory tests and test result ranges) which should be present in the case description patient. EV computed a case description score using existing thyroid disorder indices and gave advice on the suitability of the (sub)diagnosis in the patient's case. If the score was below a preset threshold, the GP was given decision support with a knowledge item list. When the GP had evaluated all case descriptions, EV presented a user mean score. The researcher compared individual user mean scores with the user population's average mean score. The experiment showed that the GPs had a reasonably well knowledge of thyroid disorders; that there were differences in scores between rural and urban GPs and GPs with different degrees of work experience; and that the indices could be applied as a basis in the prospective study. The GPs' indications led to the conclusion that EV can give decision support to GPs in daily practice and can very well be used in education and GP refresher courses.

Chapter 8 and its addendum cover the results of the prospective patient-oriented study of the development of prediction instruments for hyperthyroidism and hypothyroidism. With an instrument, a predictive score can be obtained for a patient for whom a GP wants to request a TSH test. This score can be transformed into a pretest probability. If the probability exceeds a certain probability threshold, a TSH test is adequate. The most important results were that instruments could be created for clinical/subclinical hyperthyroidism and clinical hyperthyroidism using the methods/techniques for data acquisition and analysis and the patient population described in chapter 4; that investigated items (signs/symptoms) for hyperthyroidism had relatively low raw odds ratios; that initially in the prospective study too few clinical hypothyroid patients were detected for instrument creation, and because of this data acquisition was continued concerning new clinical hyperthyroid and clinical hypothyroid patients from 1 July 1994 until 1 November 1996; that an instrument for clinical hypothyroidism could be created using the earlier mentioned methods/techniques and the cumulative population of

779 patients (87 clinical hyperthyroid, 22 subclinical hyperthyroid, 48 clinical hypothyroid, 46 subclinical hypothyroid, and 576 euthyroid patients); that the additional data did not result in a modification of the instrument for clinical hyperthyroidism; that all the hypothyroid patients experienced fatigue (fatigue is applied as a prerequisite for instrument use); that the selective data acquisition had consequences for logistic regression; that the instruments' test characteristics could be determined at various probability thresholds; and that the instruments could be evaluated with the Grouped Jackknife Method. For daily use, one threshold per instrument should be recommended in work procedures, and the instruments can be incorporated in medical DSSs. Their clinical implications and the related cuts in test-ordering expenditure led to the conclusion that the prediction instruments are capable of giving decision support about TSH test request adequacy and of stimulating a rational TSH test request behaviour at low probability thresholds.

**Chapter 9** deals with the retrospective patient-oriented study of thyroid disorder diagnoses in hospital, which was initiated because the prospective study had suggested that the GPs had diagnostically missed a number of new clinical thyroid patients. The aims were to investigate: the nondetection rate by the GPs (i.e. the hospital detection rate) in the initial data acquisition period, whether the GPs could have diagnosed the patients, and how and where the patients were diagnosed in hospital. Use was made of the laboratory information system (case finding), medical records, discharge letters, and GPs' referral letters. An internist was consulted to assess the GPs' diagnostic (in)adequacy. Results were that the GPs had a high nondetection rate; that 46% of the outpatients in Internal Medicine or Cardiology clinics, 67% of the outpatients in Neurology, Dermatology, Orthopaedics, and Throat/Nose/Ear clinics, 39% of the inpatients in the Internal Medicine clinic, and 29% of the inpatients in Cardiology, Pulmonology, Intensive Care, and Dermatology clinics had pretest probabilities of hyperthyroidism or hypothyroidism which justified TSH and FT<sub>4</sub> tests; that most patients were referred by GPs for nonthyroid diseases; and that some hospital diagnoses were fortuitous. Reasons for the nondetection of thyroid patients by the GPs (e.g. insufficiently clear clinical pictures) could not be assessed conclusively; a prospective study might be more suitable.

**Chapter 10** comprises conclusions and a general discussion of the research as well as related findings in literature. All principal research questions could be answered on the basis of research results:

- Literature provided insight into the state of the art of GP information systems and DSSs.
- Active support for checking the adequacy of a diagnosis in a patient's case using criteria belonging to that diagnosis was found to enhance the quality of registered diagnoses and medical record completeness.
- GPs proved to be more satisfied with active decision support provided by a DSS regarding the interpretation of diagnostic criteria than with the DSS's passive decision support regarding unsolicited criteria presentation.
- GPs were found to have a reasonably good knowledge of the medical domain of thyroid disorders when compared to the knowledge stored in the models of a DSS.
- Prediction instruments which indicate TSH test adequacy on the basis of signs/symptoms indeed could be developed.
- Eighty-seven patients with thyroid disease for whom GPs had not requested a TSH test were detected in hospital during a ten-month period against 45 in primary care. Of the 87 patients, 55 were detected in outpatient and 32 in inpatient clinics. It has not become clear from the retrospective study with its limited possibilities why the GPs had not considered a thyroid diagnosis in these patients before referral.

Research results and findings in literature can have consequences concerning DSS development and evaluation, data registration, guideline modification and stipulation regarding thyroid disorders, the TSH and FT<sub>4</sub> test-ordering behaviour of GPs and related cuts in test-ordering expenditure, prediction instrument creation, evaluation, and validation (particularly through a primary care pilot study), and the application and stipulation of probability thresholds. Recommendations for future research concern:

- a further evaluation of the prediction instruments in general practices,
- a prospective study of patients who are referred to hospital for nonthyroid diseases and for whom abnormal TSH and FT<sub>4</sub> outcomes are determined in hospital but for whom no TSH tests are previously requested by GPs,
- a further investigation of data entry methods for data acquisition in the form of standardized terminology required for use in DSSs, and
- an identification of other medical domains in which GPs need support on the basis of epidemiological data.

154

154

154

154

154

154

154

154

154

154

154

154

154

154

154

## Chapter 12

### Samenvatting

Met dit proefschrift wordt beoogd een wetenschappelijke verhandeling te geven van het onderzoek op het gebied van diagnostische beslissingsondersteuning voor huisartsen. InformatieTechnologie (IT) binnen beslissingsondersteuning in de eerste lijn en besluitvorming in de eerste lijn inzake schildklierziekten zijn de twee hoofdgebieden van onderzoek die behandeld worden. In detail zal het proefschrift rapporteren over het onderzoek naar:

- de geschiktheid van computergebaseerde beslissingsondersteuning in de eerste lijn, d.w.z. de toepassing van IT ter ondersteuning van het beslissingsproces van huisartsen, en
- de structurering van medische kennis op het gebied van schildklierziekten voor beslissingsondersteunende doeleinden, d.w.z. de ontwikkeling van (computergebaseerde) voorspellingsmodellen voor schildklierziekten waarmee huisartsen de toepasselijkheid van de Thyroid-Stimulating Hormone (TSH)-test in medische gevallen kan bepalen.

Achtergrondinformatie over het onderzoek, de belangrijkste onderzoeksvragen, en de opbouw van het proefschrift worden beschreven in **hoofdstuk 1**. Aanleiding tot het onderzoek was een enquête onder huisartsen die in theorie had aangetoond dat deze inderdaad behoefte hadden aan beslissingsondersteuning door BeslissingsOndersteunende Systemen (BOS'en), dat zowel actieve als passieve ondersteuning benodigd was in veel situaties, en dat het medisch gebied van schildklierziekten een probleemgebied vormde waarin beslissingsondersteuning welkom was. De belangrijkste onderzoeksvragen in het IT-onderzoek waren:

- *Wat is de stand van zaken op het gebied van huisartsinformatiesystemen en BOS'en?*
- *Kan de kwaliteit van geregistreerde diagnoses en de compleetheid van het medisch dossier worden verbeterd door BOS'en?*
- *Met welk type beslissingsondersteuning - actieve of passieve - dat door BOS'en wordt aangeboden, zijn huisartsen het meest tevreden?*
- *Hoe is het gesteld met de kennis van huisartsen inzake het medische gebied van schildklierziekten in vergelijking met de kennis die opgeslagen is in de modellen van een BOS?*

De belangrijkste onderzoeksvragen in het schildklierkennis-onderzoek waren:

- *Kunnen voorspellingsmodellen gebaseerd op klachten en symptomen worden ontwikkeld die aangeven of een TSH-test toepasselijk is?*

- *Hoeveel schildklierpatiënten voor wie huisartsen geen TSH-test aanvragen, worden gedetecteerd in het ziekenhuis; waar worden ze gedetecteerd; en waarom worden ze niet gedetecteerd in de eerste lijn?*

Terwijl de hoofdstukken 1, 2, en 10 beide onderzoeksgebieden beschrijven, wordt het IT-onderzoek behandeld in de hoofdstukken 3 en 5 tot en met 7 en brengen de hoofdstukken 4, 8, en 9 verslag uit over het schildklierkennis-onderzoek.

**Hoofdstuk 2** bevat de resultaten van een literatuurstudie, en is opgebouwd uit drie onderdelen. Het onderdeel over huisartsinformatiesystemen beschrijft bevindingen in de literatuur betreffende de evolutie in eerste-lijnssystemen, de ontwikkeling, voor- en nadelen, en vereisten voor een goed functioneren van patiëntinformatiesystemen in de eerste lijn, en de huidige situatie in Nederland betreffende het gebruik van computers en andere hulpmiddelen door huisartsen. Het onderdeel over beslissingsondersteuning voor huisartsen behandelt bevindingen in literatuur betreffende de evolutie in systeemtypen, de ontwikkeling en implementatie van BOS'en, hun evaluatie, en hun medische invloed. Het onderdeel over de medische kennis inzake schildklierziekten brengt verslag uit over bevindingen in literatuur betreffende het aanvragen van laboratoriumtesten door huisartsen en vroegere en huidige hulpmiddelen in het diagnostiseren van schildklierziekten.

In **hoofdstuk 3** worden het volledig geïntegreerde HuisartsInformatie- en OnderzoeksSysteem HIOS en BOS HIOS+ beschreven. De systemen zijn ontwikkeld met behulp van de Phased Development Methodology (PDM). HIOS bestaat uit zes modulen, heeft al de eigenschappen van een medisch registratiesysteem ter ondersteuning van huisartsen in de dagelijkse praktijkactiviteiten, en biedt faciliteiten voor het uitvoeren van onderzoek. Het werd geïmplementeerd in een huisartspraktijk en een gezondheidscentrum, en werd positief geëvalueerd. HIOS+ bestaat uit HIOS en zeven beslissingsondersteunende modulen/BOS'en. Het helpt huisartsen bij de (diagnostische) besluitvorming door het presenteren van tekstboekinformatie, het genereren van adviezen en waarschuwingen, het creëren van modellen, het opstellen van rapporten, en het evalueren van kennis voor onderwijsdoeleinden. Alle evaluaties van HIOS+-modulen door huisartsen hadden positieve resultaten. De volledige integratie van een medisch registratiesysteem met BOS'en en het gebruik van een modellenbestand los van de software biedt voordelen. De conclusie was dat HIOS en HIOS+ ondersteuning kunnen bieden in de diagnostiek en de behandeling van ziekten, en gebruikt kunnen worden voor het screenen en het toezicht houden op patiëntenpopulaties.



**Hoofdstuk 4** beschrijft de opzet van de prospectieve patiëntgebonden studie naar voorspellingsmodellen voor schildklierziekten. Gegevens werden verkregen met behulp van algemene procedures voor het aanvragen van TSH- en Free Thyroxine (FT<sub>4</sub>)-testen door huisartsen en voor het interpreteren van testresultaten, studie-specifieke procedures voor het toepassen van de inclusiecriteria en voor het hanteren van de enquêteformulieren, en gecreëerde en gevalideerde patiënt- en huisartsenquêtes en het formulier van het Diagnostisch Coördinerend Centrum (DCC). Van 1 september 1993 tot en met 1 juli 1994 werd door 80 huisartsen uit Maastricht e.o. een TSH-test aangevraagd voor 668 nieuwe patiënten (37 klinische hyper-, 21 subklinische hyper-, 8 klinische hypo-, 37 subklinische hypo-, en 565 euthyreoïdiepatiënten). Hun gegevens werden geanalyseerd betreffende gegevenskwaliteit in de zin van compleetheid en correctheid. De kwaliteit van de gegevensinvoer en de gegevensbank werd gecontroleerd met het registratiesysteem THYPRED. In de gegevensanalyse voor het creëren van voorspellingsmodellen werden de gegevens univariaat, bivariaat, en multivariaat (factoranalysen en stapsgewijze logistische-regressie-analysen) geanalyseerd. Voor het evalueren van de voorspellingsmodellen werd de Grouped Jackknife Method toegepast.

In **hoofdstuk 5** is een IT-experiment beschreven dat de bruikbaarheid van het BOS Formal Interpreter (FI) betrof in de validatie van International Classification of Primary Care (ICPC)-codes gedurende gegevensregistratie, gebruikmakend van International Classification of Health Problems in Primary Care (ICHPPC)-II-Defined-criteria in vrijetekstformaat. Met het BOS Free-text Formalizer (FF) werden voor 101 ICPC-codes voor de computer verwerkbare queries (vraagstructuren) gecreëerd, die werden gevalideerd door een onafhankelijk nationaal huisartsenpanel. In het experiment werden casussen voorgelegd aan drie huisartsengroepen van gelijke grootte, die respectievelijk geen, passieve beslissingsondersteuning (criteriapresentatie), en actieve beslissingsondersteuning (adviezen en waarschuwingen gebaseerd op het verwerken van queries met gegevens uit het medisch dossier) van FI ontvingen. Inzake de casussen moesten de huisartsen diagnosticeren en diagnose-ondersteunende gegevens registreren. Deze werden gevalideerd door een huisarts en een onafhankelijk huisartsenpanel, en werden geanalyseerd, ook betreffende het statistisch significant zijn van verschillen tussen de groepen (de Mann-Whitney test), door de onderzoeker met betrekking tot de invloed van FI. Er werd geconcludeerd dat actieve beslissingsondersteuning resulteert in een hogere diagnostische correctheid, een groter diagnostisch vertrouwen, en een completer medisch dossier; dat vrijetekstcriteria niet rechtstreeks vertaald kunnen worden in queries; en dat de Arden Syntax for Medical

Logic Modules een flexibel hulpmiddel is in de creatie van queries.

**Hoofdstuk 6** behandelt een IT-experiment voor het bepalen van de gebruikerstevredenheid met FI. In een laboratoriumstudie werden drie gelijkgroten huisartsengroepen, die respectievelijk geen beslissingsondersteuning (groep I), passieve -ondersteuning (groep II), en actieve -ondersteuning (groep III) ontvingen, gevraagd om gegevens inzake casussen te registreren. De gebruikerstevredenheid werd bepaald met een aangepast Bailey-enquêteformulier. Gemiddelde en genormaliseerde groepscores werden berekend; de gebruikers bleken 'enigszins' tevreden te zijn met het functioneren, de prestatie, de flexibiliteit, en de gebruikersbetrokkenheid van FI, alsmede de integratie van FI met HIOS. Huisartsen in groep III waren tevredener dan de anderen; huisartsen in groep II waren het minst tevreden. Een statistisch significant verschil (bepaald met de two-sample-t-test) bestond er tussen de groepen II en III, maar niet tussen de groepen I en III; het additioneel stellen van vragen aan huisartsen in groep III door FI werd blijkbaar niet als storend ervaren. Open vragen toonden aan dat de huisartsen actieve en passieve ondersteuning algemeen bruikbaar achtten, dat ze graag beide typen beslissingsondersteuning gehad zouden willen hebben, en dat ze een grote verscheidenheid aan voordelen verwachten van deze typen. In een veldstudie werd FI gedurende drie maanden door huisartsen gebruikt in een plattelandshuisartspraktijk. Gemiddelde en genormaliseerde groepscores werden bepaald, gebruikmakend van het aangepaste Bailey-enquêteformulier; de gebruikers bleken 'zeer' tevreden te zijn met FI. Op grond van het experiment werd geconcludeerd dat de zeer tevredenstellende testresultaten van FI goede hoop gaven dat medische BOS'en in de toekomst door huisartsen gebruikt zouden gaan worden.

In **hoofdstuk 7** is een IT-experiment beschreven dat het BOS Evaluator (EV) betrof en zijn functie betreffende het testen van medische kennis en het geven van beslissingsondersteuning bij kennishiaten. In het experiment werd EV gebruikt door 17 huisartsen ter bepaling van hun kennis van schildklierziekten op grond van casussen. Een huisarts moest relaties aangeven tussen een (sub)diagnose en geslacht of leeftijd, en kennis-items (b.v. klachten/symptomen, en laboratoriumtesten en hun testuitslaggrenzen) selecteren die aanwezig zouden moeten zijn in de casuspatiënt. EV berekende een casusscore, gebruikmakend van bestaande indices voor schildklierziekten, en gaf advies betreffende de toepasselijkheid van de (sub)diagnose in dit specifieke geval. Indien de score lager was dan een vooraf vastgestelde drempel, werd ondersteuning gegeven aan de huisarts in de vorm van een lijst van kennis-items. Wanneer de huisarts alle casussen had

geëvalueerd, presenteerde EV een gemiddelde score voor de gebruiker. De onderzoeker vergeleek de gemiddelde scores van alle individuele gebruikers met de gemiddelde score van de gehele populatie. Het experiment toonde aan dat de huisartsen een redelijk goede kennis hadden van schildklierziekten; dat er verschillen waren in scores tussen plattelands- en stadshuisartsen en tussen huisartsen met een verschillende mate van werkervaring; en dat de indices zouden kunnen worden gebruikt als basis in de prospectieve studie. Op grond van de indicaties van de huisartsen werd geconcludeerd dat EV in staat is beslissingsondersteuning te geven aan huisartsen in de dagelijkse praktijkvoering en dat het zeer goed gebruikt kan worden in onderwijs en opfriscursussen aan huisartsen.

**Hoofdstuk 8 en het bijbehorend addendum** brengen verslag uit van de resultaten van de prospectieve patiëntgebonden studie naar de ontwikkeling van voorspellingsmodellen voor hyper- en hypothyreoïdie. Met een voorspellingsmodel kan een voorspellende waarde worden verkregen voor een patiënt voor wie een huisarts een TSH-test wil aanvragen. Deze waarde kan worden omgezet in een pretestkans. Indien deze een bepaalde kansdrempel overschrijdt, is een TSH-test toepasselijk. De belangrijkste resultaten waren dat voorspellingsmodellen gecreëerd konden worden voor klinische/subklinische hyper- en klinische hyperthyreoïdie, gebruikmakend van de methoden/technieken voor gegevensacquisitie en -analyse en de patiëntenpopulatie beschreven in hoofdstuk 4; dat onderzochte items (klachten/symptomen) voor hyperthyreoïdie relatief lage odds ratios hadden; dat aanvangs in de prospectieve studie te weinig klinische hypothyreoïdiepatiënten waren gedetecteerd voor modelcreatie, en dat derhalve van 1 juli 1994 tot en met 1 november 1996 werd doorgegaan met de gegevensacquisitie betreffende nieuwe klinische hyper- en klinische hypothyreoïdiepatiënten; dat een voorspellingsmodel voor klinische hypothyreoïdie kon worden gecreëerd, gebruikmakend van de eerder genoemde methoden/technieken en de cumulatieve patiëntenpopulatie van 779 patiënten (87 klinische hyper-, 22 subklinische hyper-, 48 klinische hypo-, 46 subklinische hypo-, en 576 euthyreoïdiepatiënten); dat de additionele gegevens niet leidden tot aanpassing van het voorspellingsmodel voor klinische hyperthyreoïdie; dat alle hypothyreoïdiepatiënten aan vermoeidheid leden (vermoeidheid wordt toegepast als voorwaarde voor het gebruik van het voorspellingsmodel); dat testkarakteristieken van de voorspellingsmodellen konden worden bepaald bij verschillende kansdrempels; en dat de voorspellingsmodellen konden worden geëvalueerd met de Grouped Jackknife Method. Voor gebruik van de voorspellingsmodellen in de dagelijkse praktijk zal per model één kansdrempel in werkafspraken moeten worden aanbevolen, en kunnen de modellen in medische BOS'en worden ingebouwd. Op

grond van hun klinische implicaties en de samenhangende kostenbesparingen in testaanvragen werd geconcludeerd dat de voorspellingsmodellen in staat zijn beslissingsondersteuning te geven inzake de toepasselijkheid van de TSH-test en bij lage kansdrempels een rationeel TSH-testaanvraaggedrag te stimuleren.

**Hoofdstuk 9** behandelt de retrospectieve patiëntgebonden studie naar schildklierdiagnosen in het ziekenhuis die was opgestart aangezien de prospectieve studie had suggereerd dat de huisartsen een aantal nieuwe klinische schildklierpatiënten diagnostisch gezien niet hadden gedetecteerd. De doelstelling was onderzoek te verrichten naar: het aantal door de huisartsen in de initiële gegevensacquisitieperiode ongedetecteerde patiënten (d.w.z. het aantal gedetecteerden in het ziekenhuis), het aspect of de patiënten door de huisartsen hadden kunnen worden gedetecteerd, en de aspecten hoe en waar de patiënten in het ziekenhuis werden gediagnostiseerd. Hierbij werd gebruikgemaakt van het laboratoriuminformatiesysteem (case finding), medische dossiers, ontslagbrieven, en verwijsbrieven van huisartsen. Een internist werd geconsulteerd om de diagnostische (in)adequaatheid van de huisartsen te bepalen. Resultaten waren dat de huisartsen een groot aantal patiënten niet hadden gedetecteerd; dat 46% van de patiënten in de poliklinieken van Interne Geneeskunde en Cardiologie, 67% van de patiënten in de poliklinieken van Neurologie, Dermatologie, Orthopedie, en Keel/Neus/Oor, 39% van de patiënten in de kliniek van Interne Geneeskunde, 29% van de patiënten in de klinieken van Cardiologie, Pulmonologie, Intensive Care, en Dermatologie pretestkansen op hyper- of hypothyreoïdie hadden die TSH- en FT<sub>4</sub>-testen rechtvaardigden; dat de meeste patiënten door huisartsen waren verwezen vanwege andere dan schildklierziekten; en dat sommige diagnosen in het ziekenhuis toevallig waren. Redenen voor het niet detecteren van schildklierpatiënten door de huisartsen (b.v. het onvoldoende herkenbaar zijn van ziektebeelden) konden niet overtuigend worden vastgesteld; een prospectieve studie is hiertoe mogelijk beter geschikt.

**Hoofdstuk 10** omvat conclusies en een algemene discussie van het onderzoek alsmede gerelateerde bevindingen in de literatuur. Alle belangrijke onderzoeksvragen konden worden beantwoord op basis van onderzoeksresultaten:

- Literatuur gaf inzicht in de stand van zaken op het gebied van huisartsinformatiesystemen en BOS'en.
- Actieve beslissingsondersteuning voor het controleren van de toepasselijkheid van een diagnose in een bepaald geval, gebruikmakend van criteria behorend bij die

diagnose, bleek in staat te zijn de kwaliteit van geregistreerde diagnoses en de compleetheid van het medisch dossier te verbeteren.

- Huisartsen bleken tevredener te zijn met actieve beslissingsondersteuning aangeboden door een BOS betreffende het interpreteren van diagnostische criteria, dan met de passieve ondersteuning geboden door het BOS inzake het ongevraagd presenteren van criteria.
- Huisartsen bleken een tamelijk goede kennis te hebben van het medische gebied van schildklierziekten in vergelijking met de kennis die opgeslagen is in de modellen van een BOS.
- Voorspellingsmodellen, die de toepasselijkheid van een TSH-test aangeven op basis van klachten/symptomen, konden inderdaad worden ontwikkeld.
- Zevenentachtig schildklierpatiënten voor wie huisartsen geen TSH-test hadden aangevraagd, werden gedurende 10 maanden gedetecteerd in het ziekenhuis tegen 45 gedetecteerde patiënten in de eerste lijn. Van deze 87 patiënten waren er 55 gedetecteerd in poliklinieken en 32 in klinieken. Het is niet duidelijk geworden uit de retrospectieve studie met zijn beperkte mogelijkheden waarom de huisartsen geen schildklierdiagnose hadden overweegd bij deze patiënten vóór verwijzing.

Onderzoeksresultaten en bevindingen in literatuur kunnen uitwerkingen hebben betreffende de ontwikkeling en evaluatie van BOS'en, de registratie van gegevens, de aanpassing en stipulatie van richtlijnen voor schildklierziekten, het aanvraaggedrag van huisartsen inzake TSH- en FT<sub>4</sub>-testen en gerelateerde kostenbesparingen, de creatie, evaluatie, en validatie (vooral via een proefonderzoek (pilot study)) van voorspellingsmodellen, en de applicatie en stipulatie van kansdrempels. Aanbevelingen voor toekomstig onderzoek betreffen:

- een verdere evaluatie van de voorspellingsmodellen in huisartspraktijken,
- een prospectieve studie van patiënten die naar het ziekenhuis zijn verwezen op grond van andere dan schildklierziekten en voor wie afwijkende TSH- en FT<sub>4</sub>-testresultaten zijn bepaald in het ziekenhuis maar voor wie eerder door huisartsen geen TSH-testen waren aangevraagd,
- een verder onderzoek naar methoden voor gegevensinvoer teneinde gegevensacquisitie in de vorm van gestandaardiseerde terminologie te verkrijgen, wat een voorwaarde is voor het gebruik in BOS'en, en
- een identificatie van andere medische gebieden waarin ondersteuning van huisartsen op basis van epidemiologische gegevens noodzakelijk is.

**Appendices to the Thesis**  
**'Diagnostic Decision Support for General Practitioners'**

## Appendix I: The TSH test for assessing the thyroid function

### \* Aims

- To confirm or exclude hyperthyroidism and hypothyroidism on the basis of clinical symptoms.
- To control the thyroid function in case of a treatment of hyperthyroidism or hypothyroidism.

### \* Indications, a few rules of thumb

- In adult patients, hyperthyroidism and hypothyroidism will present themselves through classic clinical pictures (see the notes to the indications); monosymptomatic clinical pictures will occur much less frequent.
- In an elderly patient, hyperthyroidism often is obscured by a problem of a cardiovascular nature. Monosymptomatic forms, like muscular weakness, diarrhoea, and mental retardation, also frequently occur; the classic clinical picture is the exception rather than the rule.
- In an elderly patient, the diagnosis of hypothyroidism, also if it presents itself through its classic clinical picture, is fairly often missed due to the circumstance that changes (see the notes to the indications) occur gradually and are regarded as characteristic of old age. Moreover, an elderly patient often is not in the habit of complaining.

### \* Notes to the indications

- The diagnosis of a thyroid disorder does not pose problems if it presents itself through its classic clinical picture. The symptoms indicated by a patient, mostly of the female gender, are nervousness, irritability, heart palpitations, heat intolerance, excessive perspiration, and weight loss accompanied by a good or increased appetite. In examination, the following signs are found: the (female) patient is nervous, is sometimes agitated, can not sit still for a moment, has sweaty and warm palms of the hands, and experiences a fine tremor of the outstretched fingers. A fast pulse rate and a raised pulse pressure are also found; eye problems and a diffusely enlarged thyroid gland complement the clinical picture.
- In elderly patients with non-'classic' hyperthyroidism, the clinical picture fairly often is dominated by signs and symptoms of a cardiovascular nature: atrial fibrillation is frequently experienced; hyperthyroidism can be obscured by therapy-resistant cardiac decompensation and angina pectoris.
- Hyperthyroidism can present itself occasionally in a monosymptomatic way through muscular weakness (proximal muscle groups in the extremities), diarrhoea, oligomenorrhoea or amenorrhoea, mental retardation, and atrial fibrillation.
- The clinical picture of hypothyroidism can be very distinct. Hypothyroidism can present itself through the following signs and symptoms: fatigue (sleep a lot), cold intolerance, weight increase, mental changes (retardation of memory and of critical judgement), typical facial changes and deteriorated mimicry (all hypothyroid patients look alike), low and hoarse voice, monotone and slow speech (diagnosis over the telephone), and dry skin and hair.
- Patients with mild forms of hypothyroidism are often difficult to diagnose on the basis of minimal or specific symptoms. The possibility of hypothyroidism is considered in too few cases concerning patients with symptoms of having become less energetic, facial puffiness, etc. Especially in case of a history of a treated hyperfunction of the thyroid gland or a family history of an autoimmune illness, the clinical picture of hypothyroidism should be kept in mind.

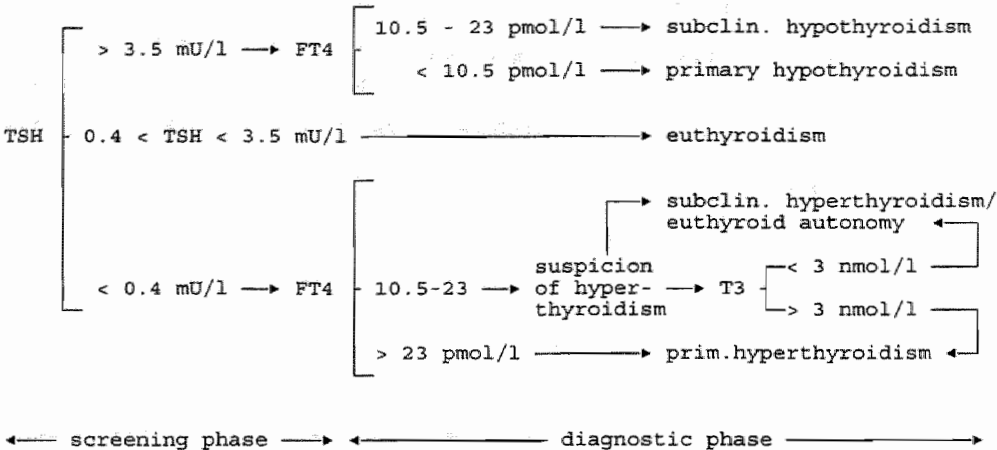
### \* Tests available for request

- The TSH test (IRMA).
- The FT<sub>4</sub> test in case of an abnormal TSH outcome.

\* Reference values and charges

Tests	Reference values	Charges (f)	
		SF	NS
FT <sub>4</sub>	10.5 - 23.0 pmol/l	45	45
TSH (IRMA)	0.4 - 3.5 mU/l	45	45

\* Interpretation



\* Notes to the interpretation

- Due to the day/night rhythm and the sensitivity of the TSH concentration, other reference values than the presented ones are used in clinical practice.
- In the outpatient Endocrinology clinic, the sensitivity, the specificity, and the predictive values of the TSH test are very high (see the table). In general practice, the positive predictive value in particular will be clearly lower due to a much lower prevalence of hyperthyroidism and hypothyroidism.

The sensitivity, the specificity, and the predictive values (PVs) of the FT<sub>4</sub> test and the TSH (IRMA) test in the diagnosis of hyperthyroidism and hypothyroidism

Tests	Reference values	Sensitivity %	Specificity %	PV+ %	PV- %
FT <sub>4</sub> pmol/l	12.1 - 23.9	69.5	84.0	56.9	90.2
TSH mU/l	0.4 - 3.9	98.9	94.3	83.9	99.7



Appendix II: The Patient Questionnaire

\*\*\*\*\*  
\* Where necessary, PLEASE complete in CAPITALS \*  
\*\*\*\*\*



-----  
Patient questionnaire

Surname : ..... ☐ male ☐ female  
Initials : .....  
Date of birth : day .... month .... year 1....  
Address : ..... No.: .....  
Place of residence : .....  
Surname of your GP : .....  
Request filing date : day .... month .... year 199..

-----

- 01 Are you lately more nervous than before?  
☐ yes  
☐ no
- 02 Are you lately more agitated in your behaviour than before?  
☐ yes  
☐ no
- 03 Are you lately more irritable than before?  
☐ yes  
☐ no
- 04 Are you lately depressed (feeling down)?  
☐ yes  
☐ no
- 05 Are you lately more forgetful than before?  
☐ yes  
☐ no
- 06 Do you have more trouble concentrating while reading or in conversations, etc. in the last few months than before?  
☐ yes  
☐ no
- 07 Are you lately more tired (after little exertion, etc.) than before?  
☐ yes  
☐ no
- 08 Do you lately perspire more or less than or the same as before?  
☐ more perspiration  
☐ less perspiration  
☐ no change
- 09 Do you lately endure heat better or less well than or equally well as before?  
☐ endure heat better  
☐ endure heat less well  
☐ no change
- 10 Do you lately endure cold better or less well than or equally well as before?  
☐ endure cold better  
☐ endure cold less well  
☐ no change
- 11 Are you lately more or less active than before or have you remained as active as before?  
☐ more active  
☐ less active  
☐ no change
- 12 Has there lately been a change in your mobility (in walking or regarding activities around the house, etc.)?  
☐ slower  
☐ agitated  
☐ no change
- 13 Have you experienced a change in appetite lately?  
☐ appetite increased

- ☐ appetite decreased
- ☐ no change

- 14 Has your weight changed lately?  
☐ increased, ..... kg in ... months  
☐ decreased, ..... kg in ... months  
☐ no change
- 15 Has your stool changed lately?  
☐ obstipation  
☐ diarrhoea  
☐ no change
- 16 Has your skin changed lately and become dryer or fattier than before?  
☐ dryer  
☐ fattier  
☐ no change
- 17 Has your skin lately become paler than before?  
☐ yes  
☐ no
- 18 Have you lately noticed tremors in your hands?  
☐ yes  
☐ no
- 19 Do you suffer from a tickling feeling in various parts of your body lately?  
☐ yes  
☐ no
- 20 Have your eyes become bulgy lately?  
☐ yes  
☐ no
- 21 Do you have red eyes lately?  
☐ yes  
☐ no
- 22 Do you have a burning sensation in your eyes lately?  
☐ yes  
☐ no
- 23 Do you have watery eyes lately?  
☐ yes  
☐ no
- 24 Do you lately experience a pain in or behind the eyes?  
☐ yes  
☐ no
- 25 Have you noticed a swelling around your eyes lately?  
☐ yes  
☐ no
- 26 Has your eyesight deteriorated (in sharpness and/or clear vision) lately?  
☐ yes  
☐ no
- 27 Do you lately suffer from double vision?  
☐ yes  
☐ no
- 28 Have you lately noticed a swelling of your throat and, if so, was the swelling painful?  
☐ yes,                      ☐ painful swelling                      ☐ painless swelling  
☐ no
- 29 Do you experience pain in the throat while swallowing?  
☐ yes  
☐ no
- 30 Has your hair lately become dryer or fattier than before or has it not changed?  
☐ dryer  
☐ fattier  
☐ no change
- 31 Do you lately suffer from hair loss?  
☐ yes

☐ no

Have you lately noticed a change in your voice and, if so, in what way did your voice change?

☐ yes, ☐ voice hoarse ☐ voice deeper ☐ voice grating ☐ .....

☐ no

Do you suffer from heart palpitations lately?

☐ yes

☐ no

If you indeed have experienced heart palpitations lately, how would you characterize them?

☐ fast and regular

☐ fast and irregular

☐ only irregular

Do you lately suffer from shortness of breath on exertion?

☐ yes

☐ no

Are you on medication?

☐ yes

☐ no

If you indeed use medication, please state the products' names and indicate how long you have been taking them?

..... ☐ shorter than 6 months ☐ longer than 6 months  
..... ☐ shorter than 6 months ☐ longer than 6 months  
..... ☐ shorter than 6 months ☐ longer than 6 months  
..... ☐ shorter than 6 months ☐ longer than 6 months  
..... ☐ shorter than 6 months ☐ longer than 6 months  
..... ☐ shorter than 6 months ☐ longer than 6 months  
..... ☐ shorter than 6 months ☐ longer than 6 months  
..... ☐ shorter than 6 months ☐ longer than 6 months

Do you smoke and, if so, since when?

☐ yes, since 19..

☐ no

If you smoke, what is your daily rate?

..... cigarettes ..... cigars ..... pipes

If you presently do not smoke, did you smoke in the past and, if so, for how long?

☐ yes, from 19.. until 19..

☐ no

Do you daily consume alcohol on weekdays?

☐ yes

☐ no

If you consume alcohol on weekdays, how many drinks do you daily have?

..... drinks

Do you consume alcohol in weekends?

☐ yes

☐ no

If you consume alcohol in weekends, how many drinks do you daily have?

..... drinks

Did you have a thyroid disorder at any time in the past?

☐ yes

☐ no

Which type of thyroid disorder, if any, did you have in the past?

☐ too fast working thyroid gland (hyperthyroidism)

☐ too slow working thyroid gland (hypothyroidism)

☐ node(s) in the thyroid gland

☐ infection of the thyroid gland

Has your past thyroid disorder, if any, been medically treated and, if so, how?

☐ yes, treatment: ☐ operation ☐ radioactive iodine ☐ tablets, tablet name: .....

☐ no

Do/did your relatives have thyroid disorders?

☐ yes

☐ no

49 If any of your relatives has or has had a thyroid disorder, please describe which relative this is/was and the type of thyroid disorder he/she is/was suffering from?  
relative thyroid disorder  
.....  
.....  
.....  
.....  
.....

50 Do you have other diseases besides a possible thyroid disorder and, if so, what are they?  
☐ yes, name of disease(s): .....  
.....  
.....  
☐ no

99999999 The next four questions need to be answered by females only 99999999

51 Do you still menstruate?  
☐ yes  
☐ no

52 If you still menstruate, does your period occur regularly (i.e. with no more than 2 days difference in the commence date of the previous menstruation)?  
☐ regular  
☐ irregular

53 On which date did your last menstruation commence?  
day .... month .... year 199..

54 If you presently do not menstruate anymore, in what year did your last period occur?  
year 19..

99999999 This concludes the section of questions only to be answered by females 999999

\*\*\*\*\*  
\* Thank you very much for your cooperation in this study \*  
\*\*\*\*\*

Appendix III: The GP Questionnaire

GP questionnaire

Patient's surname : ..... ☐ male ☐ female  
Patient's initials : .....  
Patient's date of birth : day [ ][ ] month [ ][ ] year 1[ ][ ][ ]  
Patient's address : ..... No.: [ ][ ][ ]  
Patient's place of residence : .....  
GP's surname : .....  
Request filing date : day [ ][ ] month [ ][ ] year 199[ ]



Thyroid disorder hypothesis (diagnosis GP):  
☐ hyperthyroidism ☐ hypothyroidism ☐ euthyroidism  
☐ thyroid gland not enlarged ☐ thyroid gland diffusely enlarged ☐ node  
☐ .....

Physical examination data:  
- weight : [ ][ ][ ][ ] kg  
- height : [ ][ ][ ][ ] cm  
- pulse rate : [ ][ ][ ][ ] /minute  
    ☐ regular  
    ☐ irregular  
    ☐ equal  
    ☐ not equal  
- blood pressure : [ ][ ][ ][ ] mm Hg systolic  
                  [ ][ ][ ][ ] mm Hg diastolic

Were any thyroid hormone tests requested for this patient within the past 2 years?  
☐ yes  
☐ no

If so, please provide request dates and test outcomes:

Request date	TSH test outcome	FT <sub>4</sub> test outcome	T <sub>3</sub> test outcome
[ ][ ] - [ ][ ] - 199[ ]	[ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ][ ]
[ ][ ] - [ ][ ] - 199[ ]	[ ][ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ][ ]
[ ][ ] - [ ][ ] - 199[ ]	[ ][ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ][ ]
[ ][ ] - [ ][ ] - 199[ ]	[ ][ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ]	[ ][ ][ ][ ][ ][ ][ ][ ]

Did your patient have a thyroid disorder at any time in the past?  
☐ yes  
☐ no  
If so, please state the type of disorder and the method of treatment, if any:  
☐ hypothyroidism ☐ medication  
☐ hyperthyroidism ☐ operation  
☐ toxic node ☐ radioactive iodine  
☐ atoxic struma ☐ .....  
☐ .....

Which medication does your patient take presently, if any, and since when does he/she do so?  
..... ☐ < 6 months ☐ ≥ 6 months  
..... ☐ < 6 months ☐ ≥ 6 months  
..... ☐ < 6 months ☐ ≥ 6 months  
..... ☐ < 6 months ☐ ≥ 6 months

Appendix IV: The DCC Document



Diagnostic Coordinating Centre document

Patient's surname : ..... ☐ male ☐ female  
Patient's initials : .....  
Patient's date of birth : day [ ][ ] month [ ][ ] year 1[ ][ ][ ]  
Patient's address : ..... No.: [ ][ ][ ]  
Patient's place of residence : .....  
Patient's telephone number : .....  
Patient's insurance : .....

GP's surname : .....  
Request filing date : day [ ][ ] month [ ][ ] year 199[ ]

Outcomes of requested tests:

	Request date	Outc. rec. date	Test outcome	Abnormal outcome
TSH test	.. - .. - 199.	.. - .. - 199.	[ ][ ][ ][ ][ ][ ]	<input type="checkbox"/> decreased <input type="checkbox"/> increased
FT <sub>4</sub> test	.. - .. - 199.	.. - .. - 199.	[ ][ ][ ][ ][ ][ ]	<input type="checkbox"/> decreased <input type="checkbox"/> increased
T <sub>3</sub> test	.. - .. - 199.	.. - .. - 199.	[ ][ ][ ][ ][ ][ ]	<input type="checkbox"/> decreased <input type="checkbox"/> increased

Has the patient already completed the questionnaire? (Y/N) : .  
On which date did the patient complete/hand in the questionnaire? : day .. month .. year 199.

Has the GP already completed the questionnaire? (Y/N) : .  
On which date did the GP complete/hand in the questionnaire? : day .. month .. year 199.

Where did the patient register? (Desk/Gen. practice/House call) : .

Where was the patient handed his/her questionnaire? (Desk/GP) : .

At what date and time did the patient have his/her blood sample taken? : day .. month .. year 199.  
.....

# Appendix V: Item computations based on hyperthyroid patients

Item	# items present				ORs			
	HR	EU	SHR	COM	HR	EU	SHR	COM
Nervous	27	344	13	40	1.97	0.60	1.16	1.59
Agitated	27	359	16	43	1.77	0.51	2.06	1.81
Irritable	22	386	14	36	0.79	1.00	1.13	0.91
Depressed	21	393	14	35	0.68	1.09	1.07	0.81
Forgetful	22	304	11	33	1.38	0.77	1.02	1.27
Concentration↓	20	325	13	33	0.97	0.79	1.36	1.10
Tired	33	477	14	47	2.13	0.74	0.50	1.09
Perspiration↑	25	271	10	35	2.62	0.62	1.08	1.75
Perspiration↓	1	17	1	2	1.06	0.95	1.84	1.05
Heat tolerance↑	2	59	2	4	0.45	1.16	0.99	0.70
Heat tolerance↓	22	200	8	30	2.93	0.54	1.18	2.04
Cold tolerance↑	9	41	3	12	4.07	0.35	1.88	3.09
Cold tolerance↓	7	236	6	13	0.36	1.60	0.65	0.46
More active	2	52	2	4	0.55	0.87	1.16	0.78
Less active	24	373	11	35	1.12	0.97	0.64	0.92
Reduced mobility	17	233	8	25	1.30	0.70	0.92	1.15
Increased mobility	10	141	5	15	1.22	0.95	1.03	1.15
Appetite increased	8	93	1	9	1.62	0.99	0.29	1.04
Appetite decreased	13	144	6	19	1.75	0.73	1.29	1.52
Weight increased	7	194	5	12	0.49	1.37	0.67	0.58
Weight decreased	19	143	4	23	3.50	0.60	0.72	2.02
Obstipation	6	73	2	8	1.39	0.77	0.79	1.15
Diarrhoea	9	59	3	12	2.88	0.52	1.38	2.24
Dry skin	6	223	7	13	0.32	1.50	0.87	0.50
Moist skin	3	25	2	5	1.97	0.42	2.49	2.03
Pale skin	20	178	8	28	2.78	0.50	1.38	2.03
Finger tremor	21	241	4	25	2.10	0.99	0.35	1.13
Paraesthesia	14	157	8	22	1.72	0.59	1.70	1.64
Exophthalmos	6	44	3	9	2.16	0.39	1.78	2.13
Eyes red	9	128	3	12	1.18	1.02	0.59	0.99
Burning sensation eyes	16	266	3	19	0.99	1.43	0.21	0.64
Watery eyes	10	175	3	13	0.91	1.43	0.39	0.71
Painful eyes	8	151	3	11	0.87	1.25	0.49	0.72
Puffiness around eyes	12	120	2	14	1.89	0.62	0.42	1.20
Deteriorated vision	14	239	10	24	0.90	0.82	1.38	1.03
Double vision	6	70	7	13	1.32	0.52	3.56	1.97
Swollen throat	8	104	10	18	1.26	0.41	4.32	1.93
Painful swollen throat	2	45	3	5	0.63	0.68	2.07	1.18
Pain while swallowing	9	99	6	15	1.58	0.61	2.03	1.69
Dry hair	8	152	2	10	0.87	1.48	0.34	0.63
Moist hair	3	50	5	8	0.91	0.56	3.53	1.68
Loss of hair	11	166	7	18	1.15	1.06	1.33	1.18
Voice changed	13	138	7	20	1.77	0.54	1.59	1.61
Hoarse voice	7	85	6	13	1.32	0.56	2.34	1.60
Lower voice	1	27	0	1	0.65	0.84	0	0.43
Creaky voice	2	14	0	2	2.76	0.69	0	1.27
Heart palpitations	26	301	13	39	2.33	0.63	1.58	1.94
Fast+regular palp.	10	136	5	15	1.24	0.94	1.05	1.19
Fast+irregular palp.	14	116	7	21	2.44	0.48	1.89	2.16
Only irregular palp.	2	49	1	3	0.61	1.33	0.60	0.61
Dyspnoea on exertion	30	325	13	43	3.59	0.46	1.30	2.26
Use of medication	24	325	13	37	1.57	0.74	1.36	1.47
Smoker	11	243	8	19	0.62	1.18	0.90	0.74
Alcohol use daily	4	119	3	7	0.50	1.27	0.67	0.57
Alcohol use weekends	15	292	12	27	0.71	0.98	1.39	0.94
Family thyroid history	8	116	4	12	1.18	0.88	0.97	1.10
Other diseases	11	159	6	17	1.19	0.76	1.13	1.11
Menstruation	17	311	11	28	0.83	1.41	1.06	0.89
Regular menstruation	13	217	8	21	0.97	1.15	1.11	1.01
Gender female	32	495	21	53	1.26	0.46	*	2.06
Sickfund insured	29	471	16	45	1.00	1.13	0.89	1.00
Pulse rate > 90/min	18	84	4	22	5.93	0.33	1.25	3.25
Age > 60 years	14	162	6	20	1.62	0.61	1.05	1.33

## Legend:

# items present = number of items present

\* not calculable

HR = clinical hyperthyroidism

SHR = subclinical hyperthyroidism

ORs = odds ratios

EU = euthyroidism and subclinical hypothyroidism

COM = clinical/subclinical hyperthyroidism

## Appendix VI: Item computations based on hypothyroid patients

Item	no. present		p(d/s)		ORs	
	EU	HO	EU	HO	EU	HO
Nervous	344	5	0.88	0.01	0.60	2.17
Agitated	359	6	0.88	0.02	0.51	3.46
Tired	477	8	0.89	0.02	0.74	∞
Perspiration†	271	2	0.88	0.01	0.62	0.70
Preference for heat	59	1	0.91	0.02	1.16	2.41
Preference for cold	41	0	0.77	0	0.35	0
Reduced mobility	233	7	0.88	0.03	0.70	20.33
Increased mobility	141	0	0.90	0	0.95	0
Appetite increased	93	1	0.90	0.01	0.99	1.50
Appetite decreased	144	1	0.88	0.01	0.73	0.84
Weight increased	194	5	0.92	0.02	1.37	6.72
Weight decreased	143	0	0.86	0	0.60	0
Obstipation	73	2	0.88	0.02	0.77	4.37
Diarrhoea	59	0	0.83	0	0.52	0
Dry skin	223	5	0.92	0.02	1.50	5.54
Moist skin	25	1	0.81	0.03	0.42	5.94
Pale skin	178	3	0.85	0.01	0.50	2.47
Finger tremor	241	2	0.90	0.01	0.99	0.91
Exophthalmos	44	2	0.80	0.04	0.39	6.59
Burning sensation eyes	266	5	0.92	0.02	1.43	4.07
Puffiness around the eyes	120	5	0.86	0.04	0.62	12.19
Painful swollen throat	45	2	0.87	0.04	0.68	7.63
Voice changed	138	4	0.85	0.03	0.54	5.83
Heart palpitations	301	1	0.88	0	0.63	0.25
Dyspnoea on exertion	325	5	0.87	0.01	0.46	2.44
Smoker	243	5	0.91	0.02	1.18	4.60
Alcohol use	292	5	0.90	0.02	0.98	3.21
Pulse rate > 90/min	84	0	0.79	0	0.33	0
Age > 54 years	162	5	0.87	0.03	0.61	8.05

### Legend:

no. present = number of items present  
p(d/s) = probability of a disorder given a sign/symptom  
ORs = odds ratios  
EU = euthyroidism and subclinical hypothyroidism  
HO = clinical hypothyroidism



# List of Publications

## Articles in journals

1. Dupuits FMHM, Hasman A, Heumakers TAJ, Schoonbrood GGM. A research tool for general practitioners. *International Journal of Bio-Medical Computing* 1987;21:275-286.
2. Dupuits FMHM, Winkens RAG. Medische beslissingsondersteuning geïnventariseerd. *Tijdschrift voor Medische Informatica (TMI)* 1992;21(2):79-84.
3. Dupuits FMHM. The use of the Arden Syntax for MLMs in HIOS+, a decision support system for general practitioners in The Netherlands. *Computers in Biology and Medicine* 1994;24(5):405-410.
4. Dupuits FMHM, Hasman A. User satisfaction of general practitioners with HIOS+, a medical decision support system. *Computer Methods and Programs in Biomedicine* 1995;47:183-188.
5. Dupuits FMHM, Hasman A. Evaluation of a decision support system in a medical environment. *International Journal of Bio-Medical Computing* 1995;38:155-165.
6. Dupuits FMHM. Evaluation of GPs' medical knowledge by a decision support system. *Computer Methods and Programs in Biomedicine* 1995;48(1,2):91-96.
7. Dupuits FMHM. General practitioner prediction instruments for improving test request performance regarding thyroid disorders. *Royal College of General Practitioners North & West London Faculty Magazine* Winter 1996/7;(4):8-9.
8. Dupuits FMHM. Reasons for missing thyroid disorder diagnoses in general practice. *Royal College of General Practitioners North & West London Faculty Magazine*. In press.

## Papers in proceedings and presentations

9. Dupuits FMHM, Hasman A, Heumakers TAJ, Schoonbrood GGM. Development of a GP information and research system. *Proceedings of the Seventh International Congress Medical Informatics Europe 1987 (MIE 87)*. Rome: European Federation for Medical Informatics, 1987.
10. Dupuits FMHM, Hasman A, Jansen PMCHF, Schoonbrood GGM. HIOS, een kwaliteitsondersteunend systeem. *Proceedings van het Medisch Informatica Congres 1988 (MIC 88)*. Noordwijkerhout: VMBI/MIM, 1988:293-298.
11. Dupuits FMHM, Hasman A. HIOS en besluitvormingsondersteuning. *Proceedings van het Medisch Informatica Congres 1989 (MIC 89)*. Antwerpen: VMBI/MIM, 1989:257-269.
12. Dupuits FMHM, Hasman A, Schoonbrood GGM. Decision support in a general practice. In: O'Moore R, Bengtsson S, Bryant JR, Bryden JS, eds. *Proceedings of the Medical Informatics Europe 1990 (MIE 90)*. Heidelberg: Springer-Verlag Berlin/Heidelberg, 1990:231-235.
13. Dupuits FMHM, Hasman A, Ulrichs EMJJ. ModelMaker als onderdeel binnen HIOS+. In: Hofdijk WJ, Beckers WPA. *Proceedings van het Medisch Informatica Congres 1990 (MIC 90)*. Noordwijkerhout: VMBI/MIM, 1990:431-438.
14. Dupuits FMHM, Hasman A. Designing HIOS+. In: Timmers T, Blum BI, eds. *Proceedings of the IMIA Working Conference on Software Engineering in Medical Informatics*. Amsterdam: North-Holland Elsevier Science Publishers BV, 1991:181-193.
15. Dupuits FMHM, Hasman A, Ulrichs EMJJ. Diagnostic and therapeutic value of HIOS+. In: Richards B, MacOwan H, Bryant JR, Bryant Y, Hayes G, Roberts J, eds. *Proceedings of the Current Perspectives In Healthcare Computing Conference 1991 (HC 91)*. Weybridge: The British Journal of Healthcare Computing, 1991:316-324.
16. Dupuits FMHM, Hasman A, Ulrichs EMJJ. Modelling and knowledge (re)presentation within HIOS+. In: Stefanelli M, Hasman A, Fieschi M, Talmon J, eds. *Proceedings of the Third Conference on Artificial Intelligence in Medicine Europe 1991 (AIME 91)*. Heidelberg: Springer-Verlag Berlin/Heidelberg, 1991:95-104.
17. Dupuits FMHM, Hasman A, Ulrichs EMJJ, Valkenburg JJM. Advice and warning within HIOS+. In: Adlassnig K-P, Grabner G, Bengtsson S, Hansen R, eds. *Proceedings of the Medical Informatics Europe Congress 1991 (MIE 91)*. Heidelberg: Springer-Verlag Berlin/Heidelberg, 1991:45:217-221.
18. Dupuits FMHM, Hasman A, Jansen PCMHF. Classification within HIOS+. In: Adlassnig K-P, Grabner G, Bengtsson S, Hansen R, eds. *Proceedings of the Medical Informatics Europe Congress 1991 (MIE 91)*. Heidelberg: Springer-Verlag Berlin/Heidelberg, 1991:45:749-753.
19. Dupuits FMHM, Hasman A. Beslissingsondersteuning binnen HIOS+. In: Willems JL, ed. *Proceedings van het Medisch Informatica Congres 1991 (MIC 91)*. Leuven: MIM/VMBI,

1991:223-230.

20. Dupuits FMHM, Hasman A. Supporting medical decision making. In: Richards B, MacOwan H, eds. *Proceedings of the Current Perspectives in Healthcare Computing Conference 1992 (HC 92)*. Weybridge: The British Journal of Healthcare Computing 1992:530-532.
21. Dupuits FMHM, Hasman A. HIOS+, a decision aid in medicine. In: Lun KC, Degoulet P, Piemme TE, Rienhoff O, eds. *Proceedings of the Seventh World Congress on Medical Informatics 1992 (MEDINFO 92)*. Amsterdam: North-Holland Elsevier Science Publishers BV, 1992:454-460.
22. Westerhof HP, Dupuits FMHM. What is happening in primary care in The Netherlands? In: McWilliams A, ed. *Proceedings of the Annual Conference of the Primary Health Care Specialist Group of the British Computer Society 1992 (PHCSG 92)*. Worcester: British Computer Society, 1992:21-27.
23. Dupuits FMHM, Hasman A, Jansen PMCHF, Bugter-Maessen AMA. Evaluatiemethodiek voor het HIOS+, een beslissingsondersteunend informatiesysteem. In: ten Hoopen AJ, Hofdijk WJ, Beckers WPA, eds. *Proceedings van het Medisch Informatica Congres 1992 (MIC 92)*, Rotterdam: Publicon Publishing, 1992:221-230.
24. Dupuits FMHM, Hasman A. A decision-support system to improve registration and provide better patient care. In: Richards B, MacOwan H, eds. *Proceedings of the Current Perspectives in Healthcare Computing Conference 1993 (HC 93)*. Weybridge: The British Journal of Healthcare Computing 1993:211-218.
25. Dupuits FMHM. The quality of GP referral data. In: Teasdale S, Bradley P, eds. *Proceedings of the Annual Conference of the Primary Health Care Specialist Group of the British Computer Society 1993 (PHCSG 93)*. Worcester: PHCSG, 1993:88-102.
26. Dupuits FMHM. Evaluation of a cybernetic GP system. In: Trapp R, ed. *Proceedings of the Twelfth European Meeting on Cybernetics and Systems Research 1994 (EMCSR 94)*. Vienna: World Scientific Publishing Co. Pte. Ltd., 1994:975-982.
27. Dupuits FMHM. Evaluation of GPs' medical knowledge by a decision support system. In: Barahona P, Veloso M, Bryant J, eds. *Proceedings of the Medical Informatics in Europe Congress 1994 (MIE 94)*. Lisbon: The European Federation for Medical Informatics, 1994:295-300.
28. Dupuits FMHM. Transformation of thyroid disorder knowledge into prediction instruments for general practitioners. In: Teasdale Sh, ed. *Proceedings of the Annual Conference 1995 of the Primary Health Care Specialist Group of the British Computer Society (PHCSG 95)*. Cambridge: Primary Health Care Specialist Group, 1995:84-91.
29. Dupuits FMHM. Prediction instruments for general practitioners concerning hyperthyroidism and hypothyroidism. In: Dellaportas C, ed. *Proceedings of the One Day Research Workshop of the Royal College of General Practitioners 1995*. London: Royal College of General Practitioners, 1995: 3-4.
30. Dupuits FMHM. The predictability of thyroid disorders by GPs. In: Richards B, de Glanville H, eds. *Proceedings of the Current Perspectives in Healthcare Computing Conference 1996 (HC 96)*. Swindon: The British Journal of Healthcare Computing, 1996:179-186.
31. Dupuits FMHM, Hermesen JH, Oosterhuis WP, Winkens RAG. BODE, een transmurale laboratoriumaanvraag informatiesysteem. *Proceedings van het Electronisch Medisch Dossier Congres 1997 (EMD 97)*. In press.
32. Dupuits FMHM, Pop P, Hasman A, Schouten HJA. Voorspellingsmodellen betreffende schildklierandoeningen voor de eerste lijn. *Proceedings van het Medisch Informatica Congres 1997 (MIC 97)*. In press.

### **Abstracts in journals or proceedings**

33. Dupuits FMHM, Hasman A. HIOS improves medical care. Abstracts of the 12th WONCA World Conference on Family Medicine. *The Family Physician* 1989;17(1):20.
34. Dupuits FMHM, Hasman A, Schoonbrood GGM, Ulrichts EMJJ. HIOS+, a decision support system in health care. *Proceedings of the 8th Nordic Meeting on Medical and Biological Engineering 1990 of the International Federation for Medical and Biological Engineering (IFMBE 90)*. Aalborg: Aalborg University, 1990: Abstract no. 81.
35. Dupuits FMHM. The use of the Arden Syntax for MLMs in HIOS+, a decision support system for general practitioners in The Netherlands. *Artificial Intelligence (Key Abstracts)* 1995;2:23.
36. Dupuits FMHM. Prediction instruments for thyroid disorders. *Proceedings van de Onderzoekschooldag 1997*. Nijmegen: Onderzoekschool CaRe, 1997:58-59.

## Curriculum Vitae

François Marie Hubert Marcel Dupuits was born in Meerssen, The Netherlands, on December 13<sup>th</sup>, 1959. He completed his secondary education on Atheneum beta level at the St. Maartenscollege in Maastricht in 1979. Subsequently, he studied mathematics and physics at the Catholic University of Nijmegen for two years. In 1981, he transferred to the Polytechnic Heerlen (HTS-Heerlen) to study informatics. Here, he obtained a B.Sc. degree in this field in 1985. That year, François became employed as an informatician with the Department of Medical Informatics of the Maastricht University, which at that time was known as the University of Limburg. Since then, he has been carrying out scientific research, and has been taking part in the education of students. His educational duties involve lecturing various facets of informatics to students of medicine and health science, and to supervise students of informatics during their vocational training. In his research, he focuses on the development and the application of information systems, decision support systems, and other decision tools for General Practitioners (GPs). In 1989, his research endeavours culminated in the GP Information and Research System HIOS, which has all the characteristics of a medical registration system and additionally offers research facilities. Subsequently, he extended his research interest to the development and the employment of GP decision support systems. He fully integrated several decision support systems with HIOS into a HIOS+ environment. From 1993 onwards, he has been performing patient-oriented research with the aim to create thyroid disorder prediction instruments for GPs. In this research, he works closely together with the Diagnostic Coordinating Centre (DCC) and several departments of the University Hospital Maastricht. In the line of his research, François has been project manager of various projects. In the nationwide 'International Classification of Primary Care (ICPC) Thesaurus Project' (1989-1992), which was funded by the Ministry of Public Health, Welfare, and Sports (VWS), for example, he had the entire supervision over all managerial aspects of the project, including the financial administration. François has published his research work in many articles and papers in national and international journals, and has presented his research results at various national and international conferences. This thesis is the result of François' research into both decision support systems and thyroid disorder prediction instruments for GPs. The research was performed at the Department of Medical Informatics within an organizational affiliation with the Research Institute for Extramural and Transmural Health Care (Institute ExTra) of the Netherlands School of Primary Care Research (Graduate School CaRe).

## List of Abbreviations and Acronyms

ASCII	American Standard Code for Information Interchange
AW	Advice and Warning
azM	University Hospital Maastricht
BMR	Basal Metabolic Rate
BOS	BeslissingsOndersteunend Systeem
BSP	Business Systems Planning
CI	Confidence Interval
CPR	Computer-based Patient Record
CPU	Central Processing Unit
DCC	Diagnostic Coordinating Centre
DFD	Data Flow Diagram
DS	Decision Support
DSS	Decision Support System
DSSH	Decision Support System in Healthcare
EDP	Electronic Data Processing
ES	Expert System
EV	EValuator
FF	Free-text Formalizer
FI	Formal Interpreter
FNeg	False Negative
FPos	False Positive
FT <sub>3</sub>	Free Triiodothyronine
FT <sub>4</sub>	Free Thyroxine
FT <sub>4</sub> I	Free Thyroxine Index
GP	General Practitioner
HAG	Department of Family Medicine
HIOS	GP Information and Research System
HIOS+	GP Information and Research System with decision support facilities
<sup>131</sup> I	radioactive iodine
ICHPPC-II-Defined	International Classification of Health Problems in Primary Care-II-Defined
ICPC	International Classification of Primary Care
ICU	Intensive Care Unit
IF	Information Facility
ISAC	Information Systems work and Analysis of Change
IT	Information Technology
KS	Knowledge Source

LR	Likelihood Ratio
MDSS	Medical Decision Support System
MI	Department of Medical Informatics
MIS	Management Information System
MLM	Medical Logic Module
MM	Model Maker
MNS	MultiNodular Struma
MU	Maastricht University
NHG	Dutch College of General Practitioners
OCO	physical examination code
OR	Odds Ratio
PC	Personal Computer
PDM	Phased Development Methodology
PDRD	Physician's Desk Reference and Database
PPR	Paper-based Patient Record
PV	Predictive Value
RG	Report Generator
RNH	Registration Network General Practices
ROC	Receiver Operating Characteristic
SDM	System Development Method
SOAP	Subjective findings, Objective findings, Assessment, and Plan
SPSS	Statistical Package for Social Sciences
T <sub>3</sub>	Triiodothyronine
T <sub>4</sub>	Thyroxine
THYPRED	THYroid PREDiction information system
TNE	Throat/Nose/Ear
TPS	Transaction Processing System
TRH	Thyrotrophin-Releasing Hormone
TSH	Thyroid-Stimulating Hormone
TSH ICMA	TSH ImmunoChemiluminescent Metric Assay
TSH IFMA	TSH ImmunoFluoroMetric Assay
TSH IRMA	TSH ImmunoRadioMetric Assay
TSH RIA	TSH Radio-ImmunoAssay
VWS	Ministry of Public Health, Welfare, and Sports
WHO	World Health Organization